

terminàlia

REVISTA SEMESTRAL
NÚM. 27 · JUNY 2023
ISSN: 2013-6692 (impresa);
2013-6706 (electrònica)

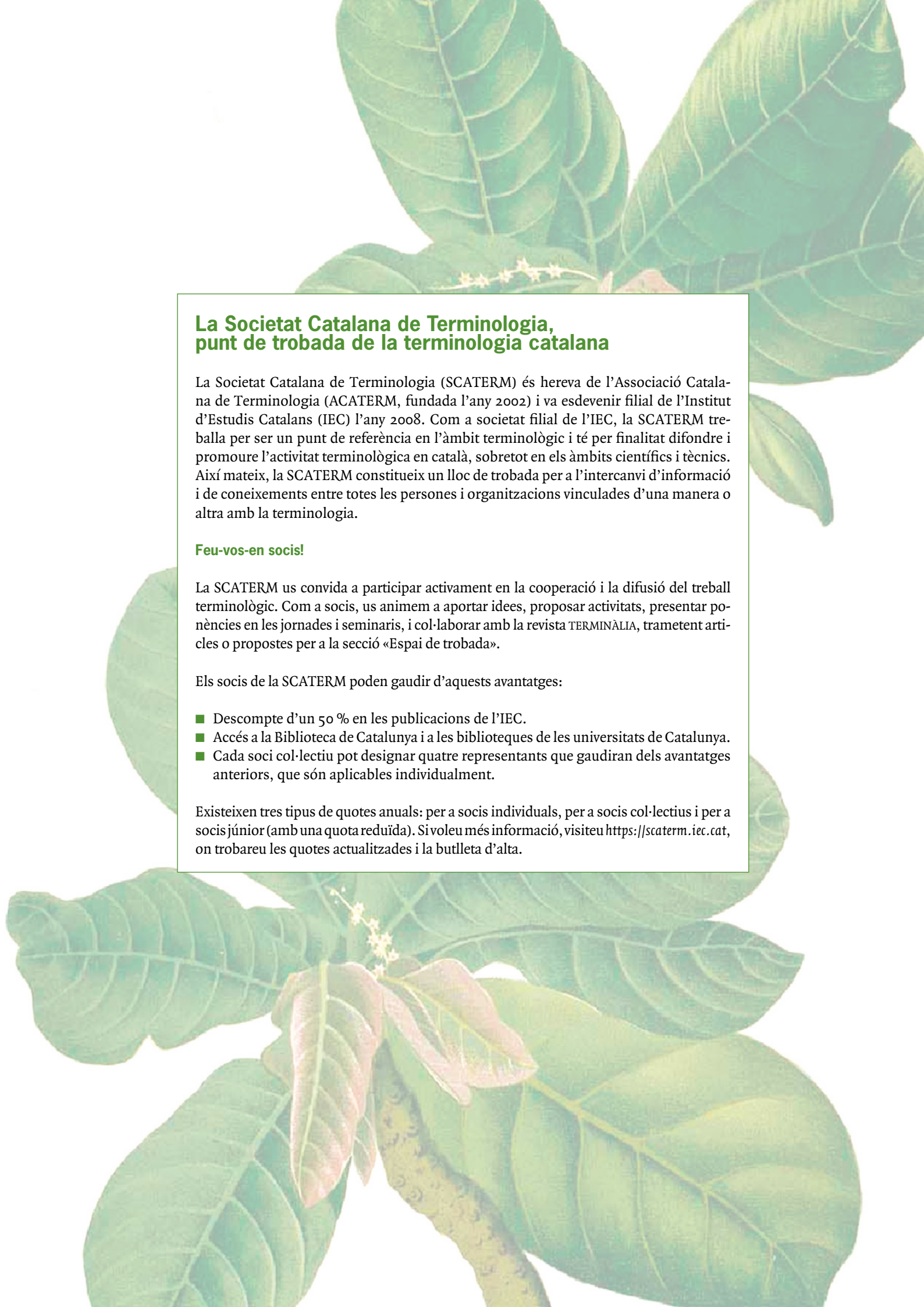


f. BOT Gènere de plantes arbòries o arbustives erectes de la família de les combretàcies, de flors petites i en general verdes o blanques, que comprèn unes 225 espècies pròpies de regions tropicals. Se n'utilitza la fusta, i també serveix com a planta ornamental i és font de tanins.

f. pl. MIT Festes en honor del déu de la mitologia clàssica romana Tèrminus, déu protector dels límits dels camps i de les fronteres. Els pagesos les celebraven anualment el 23 de febrer.

f. SCATERM Revista de la Societat Catalana de Terminologia.





La Societat Catalana de Terminologia, punt de trobada de la terminologia catalana

La Societat Catalana de Terminologia (SCATERM) és hereva de l'Associació Catalana de Terminologia (ACATERM, fundada l'any 2002) i va esdevenir filial de l'Institut d'Estudis Catalans (IEC) l'any 2008. Com a societat filial de l'IEC, la SCATERM treballa per ser un punt de referència en l'àmbit terminològic i té per finalitat difondre i promoure l'activitat terminològica en català, sobretot en els àmbits científics i tècnics. Així mateix, la SCATERM constitueix un lloc de trobada per a l'intercanvi d'informació i de coneixements entre totes les persones i organitzacions vinculades d'una manera o altra amb la terminologia.

Feu-vos-en socis!

La SCATERM us convida a participar activament en la cooperació i la difusió del treball terminològic. Com a socis, us animem a aportar idees, proposar activitats, presentar ponències en les jornades i seminaris, i col·laborar amb la revista TERMINÀLIA, trametent articles o propostes per a la secció «Espai de trobada».

Els socis de la SCATERM poden gaudir d'aquests avantatges:

- Descompte d'un 50 % en les publicacions de l'IEC.
- Accés a la Biblioteca de Catalunya i a les biblioteques de les universitats de Catalunya.
- Cada soci col·lectiu pot designar quatre representants que gaudiran dels avantatges anteriors, que són aplicables individualment.

Existeixen tres tipus de quotes anuals: per a socis individuals, per a socis col·lectius i per a socis júnior (amb una quota reduïda). Si voleu més informació, visiteu <https://scaterm.iec.cat>, on trobareu les quotes actualitzades i la butlleta d'alta.

SUMARI

EDITORIAL

Aquest número	5
---------------	---

ARTICLES

Adriana Gil Puig L'elaboració d'un diccionari d'incendis forestals adreçat a personal expert: estratègies metodològiques	7
Laura Gabriela Linares Colmenares, Amaranta Ramírez-Terrazo i Marisela Colín Rodea <i>Diccionario gastronómico de hongos mexicanos: la relevancia de un diccionario gastronómico cultural y la variación terminológica</i>	20
Julia Elena Pardo Una mirada dialectològica al Observatorio de Palabras de la RAE via Twitter	32
Sangzhu Zhaxi Descripción y clasificación: las unidades terminológicas metafóricas en los mercados de flores cortadas	51

ENTREVISTA

TERMINÀLIA parla amb... Ramon López de Mántaras i Badia, director de l'Institut de Recerca en Intel·ligència Artificial	68
--	----

DOSSIER: La terminologia de la intel·ligència artificial

Antoni Hernández Fernández i Ramon Ferrer i Cancho Lingüística quantitativa i lleis lingüístiques: de la lingüística a la intel·ligència artificial i la tecnoètica	72
Marta Villegas Montserrat El projecte AINA, la IA i les tecnologies del llenguatge	80
Ona de Gibert Bonet Resiliència digital: aprofitar el poder del col·lectiu per preservar la llengua. Una història d'èxit per al català	85

ESPAI DE TROBADA

Crònica de la XX Jornada de la SCATERM

Joan Puigmalet

90

Les Tertúlies Terminològiques de RITerm: un espai de debat sobre terminologia

Ioana Cornea

92

Recull de les activitats de la Societat Catalana de Terminologia

Societat Catalana de Terminologia

94

SEMBLANÇA:

Juan Carlos Sager (1929-2021)

Juan Carlos Sager: home polifacètic, avançat al seu temps

María Pozzi i Gerardo Sierra

97

TERMINÀLIA

Número 27 (2023)

IMATGE PORTADA:

Imatge obtinguda a partir del programa d'intel·ligència artificial Midjourney

DIRECTORA:

M. Teresa Cabré Castellví (UPF, IEC)

EDITORIA EN CAP:

Mercè Lorente Casafont (UPF, IEC)

CONSELL DE REDACCIÓ:

Ester Bonet Solé (SCATERM), Marisa Carrió Pastor (Universitat Politècnica de València), Àngels Egea Puigventós (Universitat de Barcelona), Sabela Fernández-Silva (UNED, PUCV), Glòria Fontova Hugas (TERMCAT), Anna Joan Casademont (Université TÉLUQ, Québec, Canadà), Albert Morales Moreno (Universitat Oberta de Catalunya), Pilar Sánchez-Gijón (Universitat Autònoma de Barcelona), Laura Santamaria Guinot (Universitat Autònoma de Barcelona), Johannes Schnitzer (Wirtschaftsuniversität Wien, Àustria), Mercedes Suárez de la Torre (Universidad Autónoma de Manizales, Colòmbia)

CONSELL ASSESSOR:

Gerhard Budin (Univ. de Viena), Tina Célestin (OQLF), Guiomar Ciapusio (UBA), Loïc Depecker (Univ. de París I), Pamela Faber (UGR), Joan Martí Castell (URV, IEC)

CONSELL CIENTÍFIC: Andreina Adelstein, Ieda Alves, Jordi Bover, Margarita Correia, Maria Josep Cuenca, Patrick Drouin, Joaquín García Palacios, Christophe Gerard, Xosé María Gómez Clemente, Gloria Guerrero Ramos, Mar Massanell Mesalles, Fernando Prieto Ramos, Gabriel Quiroz Herrera, Antoni Roca Rosell, Micaela Rossi, Gerardo Sierra Martínez, Joan Vallès Xirau, Lieve Vangehuchten, Carmen Varo Varo, Igone Zabala.

CAPÇALERA DE LA REVISTA: Jesús Carrasco
DISSENY I MAQUETACIÓ: Esteva&Estêvão
IMPRESSIÓ: La Gran Impressió

EN AQUEST NÚMERO HAN COL·LABORAT:
Marisela Colín, Ioana Cornea, Ona de Gibert, Ramon Ferrer, Adriana Gil, Antoni Hernández, Laura Gabriela Linares, Ramon López de Mántaras, Víctor Pàmies, Julia Elena Pardo, María Pozzi, Joan Puigmalet, Amaranta Ramírez-Terrazo, Gerardo Sierra, Marta Villegas, Sangzhu Zhaxi

SOCIETAT CATALANA DE TERMINOLOGIA

<https://scaterm.llocs.iec.cat>

Presidenta: Ester Bonet

Vicepresidenta: Rosa Estopà

Tresorer: Francesc Galera

Secretària: Teresa Tort

Vocals: Joan Puigmalet i Mireia Trias

Delegada de l'IEC: Mercè Lorente

A la plataforma de la revista podeu consultar les polítiques editorials actualitzades (<https://revistes.iec.cat/index.php/Terminalia/index>). També trobareu totes les indicacions per a les noves trameses: <https://revistes.iec.cat/index.php/Terminalia/about/submissions>



Llicència completa:

<http://creativecommons.org/licenses/by-nc-nd/3.0/es/deed.ca>

Les revistes de l'IEC allotjades a l'Hemeroteca Científica Catalana utilitzen com a descriptors les quinze propietats recomanades al Dublin Core Metadata Element Set, versió reduïda de la norma ISO 15836 (2009).

PREU D'AQUEST EXEMPLAR: 13 euros

Revista TERMINÀLIA

Societat Catalana de Terminologia
Carrer del Carme, 47 · 08001 Barcelona
Telèfon: 935 529 104
Adreça electrònica: terminalia@iec.cat

ISSN: 2013-6692 (edició impresa);

2013-6706 (edició electrònica)

Dipòsit legal: B-44926-2009

Any d'inici de la revista: 2009

Aquesta revista és accessible en línia des de:
<https://revistes.iec.cat/>, <http://terminalia.iec.cat>,
<https://issuu.com/institut-destudis-catalans>

Revista indexada a: CARHUS Plus+, CIRC, CiteFactor, DIALNET, DICE, DOAJ, Dulcinea, EBSCO, ERIH PLUS, Google Acadèmic, Hispana, IEC, ISOC, JournalTOCs, Latindex, MIAR, MLA, RACO, REDIB, RESH, RETI, SHERPA RoMEO, TIB.eu, TRACES, Ulrich's PDDb, WorldCat.

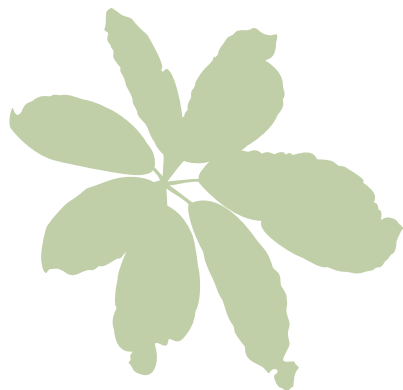
El codi ètic de la revista TERMINÀLIA inclou les recomanacions de la darrera revisió (març de 2011) del Code of Conduct and Best Practice Guidelines for Journal Editors del Committee on Publication Ethics (COPE) <https://publicationethics.org> i les impulsades des de l'Institut d'Estudis Catalans, institució editora de la revista.

Amb la col·laboració de:



Imprès en paper 100 % reciclat

Aquest número



Teniu davant vostre el número 27 de la revista TERMINÀLIA. Com recordeu, en els números semestrals de primavera les monografies tracten de temes especialitzats sectorials. Feia temps que parlàvem de dedicar un número a la terminologia de la intel·ligència artificial, però el desenvolupament de les tecnologies va ben de pressa i les aplicacions d'intel·ligència artificial generativa ja estan a l'abast de tothom. Per això hem adreçat l'«Entrevista» a Ramon López de Mántaras i Badia, un dels pioners de la intel·ligència artificial al nostre país, investigador en actiu de referència en aquest àmbit i membre de l'Institut d'Estudis Catalans. En el «Dossier» s'inclouen tres articles: el primer, d'Antoni Hernández i Ramon Ferrer, de la Universitat Politècnica de Catalunya, sobre tècniques quantitatives per caminar des de la lingüística a la intel·ligència artificial; el segon, de Marta Villegas, del Barcelona Supercomputing Center, sobre el projecte AINA, i el tercer, d'Ona de Gibert, de la Universitat d'Hèlsinki, sobre la intel·ligència artificial aplicada a les llengües minoritzades. La «Semblança» d'aquest número recorda una persona clau en el desenvolupament de les tecnologies del llenguatge, Juan Carlos Sager (1929-2021), tot un referent internacional en traducció especialitzada i terminologia, que va fundar i dirigir el Centre de Lingüística Computacional de la Universitat de Manchester. Va mantenir una relació estreta amb la terminologia catalana, durant molts anys, a través de Teresa Cabré, i va ser un dels motors per a la renovació teòrica de la terminologia. Han redactat el text de record María Pozzi, d'El Colegio de México, i Gerardo Sierra, de la Universitat Nacional Autònoma de Mèxic, que van formar-se amb Juan Carlos Sager a Manchester i amb qui van mantenir una amistat duradora.

A la secció «Articles» hem inclòs quatre textos avaluats externament per revisors mitjançant una doble revisió cega: el primer, en llengua catalana, sobre la terminologia dels incendis, a càrrec d'Adriana Gil, professional de l'àmbit que va defensar la tesi doctoral sobre el tema a la Universitat Politècnica de València; el segon, en castellà, sobre un diccionari gastronòmic de bolets de Mèxic, a càrrec de Laura Gabriela Linares



i de les seves professores Amaranta Ramírez-Terrazo i Marisela Colín, de la Universitat Nacional Autònoma de Mèxic; el tercer, també en castellà, sobre l'Observatorio de Palabras de la Real Academia Española, un recurs de difusió de novetats lèxiques a les xarxes socials, a càrrec de Julia Elena Pardo, derivat d'un treball acadèmic realitzat a la Universitat Pompeu Fabra, i el quart, en la mateixa llengua, sobre la metàfora en la terminologia de les flors tallades en xinès i en castellà, a càrrec de Sangzhu Zhaxi, de la Universitat Pompeu Fabra. Cal destacar d'aquesta nova fornada d'articles l'empenta que suposen les noves generacions per a la recerca en terminologia: dos dels articles estan vinculats a sengles tesis doctorals i els altres dos, a treballs de fi de màster.

A la secció «Espai de trobada» trobareu les cròniques de la XX Jornada de la Societat Catalana de Terminologia (SCATERM) «La terminologia en l'educació, entre la innovació i l'abús», que s'ha celebrat el 31 de maig de 2023. Complementa la secció un text sobre

la iniciativa de les Tertúlies Terminològiques, posades en marxa per RITerm, la Xarxa Iberoamericana de Terminologia, a càrrec de Ioana Cornea, professora de la Universitat Nacional Autònoma de Mèxic i membre del Comitè Executiu actual de la xarxa, que també anuncia el simposi que se celebrarà a Natal (Brasil) el proper mes de setembre. Per tancar la secció recordem les vint primeres jornades científiques organitzades per la SCATERM. Per molts anys!

Com heu pogut comprovar, l'Institut d'Estudis Catalans ha actualitzat la versió de la plataforma Open Journal System per a la publicació en línia de les revistes en obert. De mica en mica, hi anirem incorporant altres informacions i noves millores. Si hi detecteu cap error, us agraïrem que ens ho feu saber. I gràcies, també, per ajudar-nos a difondre la revista i els continguts d'aquest número. Bon estiu a tothom!

EL CONSELL DE REDACCIÓ



TERMINÀLIA és la revista de la Societat Catalana de Terminologia, SCATERM, filial de l'Institut d'Estudis Catalans. Té una periodicitat semestral i es publica en paper i en suport electrònic.

Són ben reconegudes arreu les activitats terminològiques catalanes (en formació, en recerca, en normalització, en producció de recursos). **TERMINÀLIA** és, doncs, una publicació catalana amb una clara voluntat de projecció exterior; per aquest motiu, les llengües de la revista són el català i l'anglès. A la secció «Articles» s'admeten també originals en castellà i en francès. A la resta de seccions, quan hi ha contribucions en llengües diferents de les quatre esmentades, els textos es tradueixen al català per a la versió impresa de la revista i, a la versió electrònica, es publica tant l'article en la llengua original com la traducció al català.

La temàtica de la revista gira entorn de la terminologia, en sentit ampli, i també s'estén a àmbits afins (anàlisi del discurs especialitzat, traducció especialitzada, història del discurs científic, ensenyament de llengües amb propòsits específics, enginyeria lingüística aplicada al treball terminològic, corpus lingüístics de contingut especialitzat, etc.).

El 100 % dels continguts de la revista són contribucions originals. La revista està estructurada en seccions fixes, que pretenen reflectir la situació actual de les diverses activitats terminològiques i d'àmbits afins, i també servir d'òrgan d'intercanvi i d'expressió dels diversos perfils professionals vinculats a la terminologia. És un objectiu principal de la revista publicar articles de qualitat i que hagin superat una doble revisió cega.

L'elaboració d'un diccionari d'incendis forestals adreçat a personal expert: estratègies metodològiques

ADRIANA GIL PUIG

Universitat Politècnica de València
ORCID: 0009-0005-6382-0043
a.gilpuig@edu.gva.es

Adriana Gil Puig és enginyera de

Forests i doctora en Lingüística Aplicada per la Universitat Politècnica de València, on va defensar la tesi titulada *La terminologia catalana dels incendis forestals. Recerca, anàlisi i proposta de diccionari especialitzat català-castellà-anglès* (2022). Ha exercit com a tècnica d'incendis del Ministeri d'Agricultura, Alimentació i Medi Ambient (MAGRAMA) a Castella i Lleó, Aragó i el País Valencià. Actualment, és professora de l'àrea STEAM (ciència, tecnologia, enginyeria, art i matemàtiques) a la Generalitat Valenciana. La seua línia d'investigació principal és la terminologia forestal catalana. Compta amb diverses publicacions, d'entre les quals destaca l'obra *Vocabulari forestal* (Xarxa Vives d'Universitats, 2010).



Resum

El foc forestal és un element consubstancial amb la Mediterrània que, arran del canvi climàtic i de l'abandonament rural, ha esdevingut una problemàtica ambiental de primer ordre. Presentem el treball d'elaboració d'una aplicació terminològica sobre incendis forestals adreçada a especialistes que desenvolupen la seva activitat professional en l'àmbit lingüístic català. Expliquem succintament el procés de confecció del diccionari especialitzat posant l'èmfasi en la importància de delimitar el tema amb coneixença i de tenir en compte les necessitats professionals a l'hora de dissenyar-lo perquè siga útil socialment.

PARAULES CLAU: diccionari especialitzat; terminologia catalana; terminologia forestal; especialistes; incendis forestals

Abstract

Compiling a forest fire dictionary for experts: Methodological strategies

Forest fires are an inseparable aspect of the Mediterranean region but they have become a major environmental problem because of climate change and rural abandonment. This paper describes the process of developing a terminological application on forest fires addressed to experts working in the Catalan-speaking territories. The author briefly explain the method of compiling this specialized dictionary and emphasize the importance of delimiting its scope on the basis of a thorough knowledge of the subject while taking professional needs into consideration in order to ensure its social utility.

KEYWORDS: specialized dictionary; Catalan terminology; forestry terminology; experts; forest fires

TERMINÀLIA 27 (2023): 7-19 · DOI: 10.2436/20.2503.01.185
Data de recepció: 29/03/2023. Data d'acceptació: 30/04/2023
ISSN: 2013-6692 (impresa); 2013-6706 (electrònica) · <http://terminalia.iec.cat>

1 Introducció

Els incendis forestals han esdevingut un tema candent a la Mediterrània a conseqüència del canvi climàtic, del despoblament rural i de l'abandonament de les pràctiques agrosilvoculturals. Els nostres boscos avui són molt més vulnerables perquè el paisatge mosaic que exercia de tallafoc natural s'ha perdut, la quantitat de biomassa s'ha incrementat i la vegetació presenta uns nivells d'estrès inèdits a causa de la pujada de les temperatures generalitzada i de l'augment de la recurrència d'episodis extrems com onades de calor, tempestes seques o períodes prolongats de sequera (Moreno *et al.*, 2015).

Als territoris de parla catalana estem avesats als focs forestals perquè són un fenomen ben antic i arrelat al territori (Pausas, 2012), però totes les dades apunten que estem experimentant un canvi en el règim d'incendis (Andrade *et al.*, 2019). En el segle XXI, els índexs de perill continuaran augmentant d'una manera progressiva: els incendis forestals seran cada vegada més freqüents i més severos (Moreno *et al.*, 2005). Els grans incendis forestals (GIF) d'avui no només es propaguen amb unes velocitats extremament ràpides i amb unes intensitats tèrmiques excepcionals, sinó que sovint també ho fan de manera simultània, sobre combustibles vegetals extraordinàriament secs i inflamables i en zones d'interfície urbanoforestal. Alguns incendis, com els de convecció, són capaços d'alliberar tanta energia que poden alterar les condicions atmosfèriques i generar tempestes de foc i unes dinàmiques imprevisibles. Hem après que, en unes condicions determinades, alguns grans incendis resulten senzillament inabastables i queden fora de la capacitat d'extinció, per molts mitjans humans i materials que tinguem per combatre'ls (Costa *et al.*, 2011).

El comportament del foc està canviant i l'única solució és adaptar-s'hi, aprendre a conviure-hi, prendre consciència del risc, de la incertesa i adquirir la capacitat de gestionar-la. Cal canviar el vell paradigma de l'extinció total per un sistema d'extinció amb pensament estratègic, basat en les lliçons apreses, que integre incendis de baixa intensitat que puguin crear discontinuïtats, i, sobretot, un sistema basat en la gestió forestal i en una economia rural que promogui la resiliència sistèmica (Castellnou *et al.*, 2003).

Analitzar el règim d'incendis al nostre territori és una tasca complexa perquè implica dominar molts conceptes, sovint interrelacionats. Els incendis forestals són un camp força interdisciplinari, en què intersequen matèries diverses com la meteorologia, l'ecologia, la botànica, la geomorfologia, l'enginyeria forestal, la física o la química, entre d'altres; entendre'l requereix una expertesa que s'adquireix amb molta formació. Hi ha tot un col·lectiu professional dedicat, a casa nostra, a lluitar contra aquests grans incendis, a analitzar i gestionar el foc. Com en altres disciplines científiques i tècniques, el coneixement especialitzat sobre el

foc forestal es transmet entre especialistes mitjançant paraules que tenen un significat precís: els termes.

La terminologia del foc forestal és vasta i complexa, fortament arrelada al territori, vertebrada entorn de la metàfora bèl·lica, a cavall entre la ciència i la tècnica, entre la saviesa rural popular i les noves tecnologies, entre el tecnicisme i la divulgació en els mitjans de comunicació, de vegades, opaca i densa, i d'altres, força transparent. Siga com siga, la riquesa de la terminologia d'aquest àmbit és extraordinària.

En aquest article presentem el projecte d'elaboració d'un diccionari d'incendis forestals adreçat a personal expert de l'àrea del domini lingüístic català (Gil, 2022), posant l'èmfasi en dues qüestions que són essencials perquè una obra terminològica siga reeixida: delimitar bé el camp de treball i dissenyar la proposta atenent a les necessitats de les persones destinatàries. Volem fer notar que l'elaboració d'aquest recull terminològic parteix de la motivació principal que a l'Estat espanyol no hi ha cap obra d'aquestes característiques i que el que pretén és contribuir a normalitzar —i prestigiar— l'ús del català en aquest àmbit professional.

2 Metodologia

Des del punt de vista teòric, el nostre treball se centra en l'estudi de la terminologia *in vivo* d'un camp d'especialitat i és de base descriptiva i comunicacional (Cabrè, 1999). Tot el procés d'elaboració del diccionari ha estat regit per dos principis metodològics: qualitat i adequació (Estopà, 2013); per això, des del començ, les necessitats i les preferències de les persones destinatàries han estat una qüestió central a l'hora de dissenyar tant la macroestructura com la microestructura del diccionari.

Per realitzar el diccionari especialitzat sobre incendis forestals vam seguir la metodologia següent (vegeu la figura 1).

En primer lloc, vam dur a terme una exploració de les necessitats terminològiques del personal expert en incendis forestals. Volíem esbrinar quin era el perfil —o perfils— de l'especialista sobre incendis forestals que hi havia, quin era el seu context sociolingüístic professional, els diccionaris sobre incendis que manejava i els usos que en feia o en volia fer, la motivació amb què consultava diccionaris i si aquests responien a les seves expectatives. Paral·lelament, vam fer una anàlisi sobre els reculls terminològics existents sobre incendis forestals —tant en català com en castellà i en anglès—, per conèixer les mancances i les fortaleces que presentaven, així com les necessitats del personal expert que podien satisfer i les que quedaven per resoldre.

En segon lloc, vam dissenyar la recerca terminològica. Partint dels resultats de l'exploració prèvia, vam prendre les decisions següents:

- Tipus de recerca: sistemàtica descriptiva.
- Dimensions: un miler de termes.

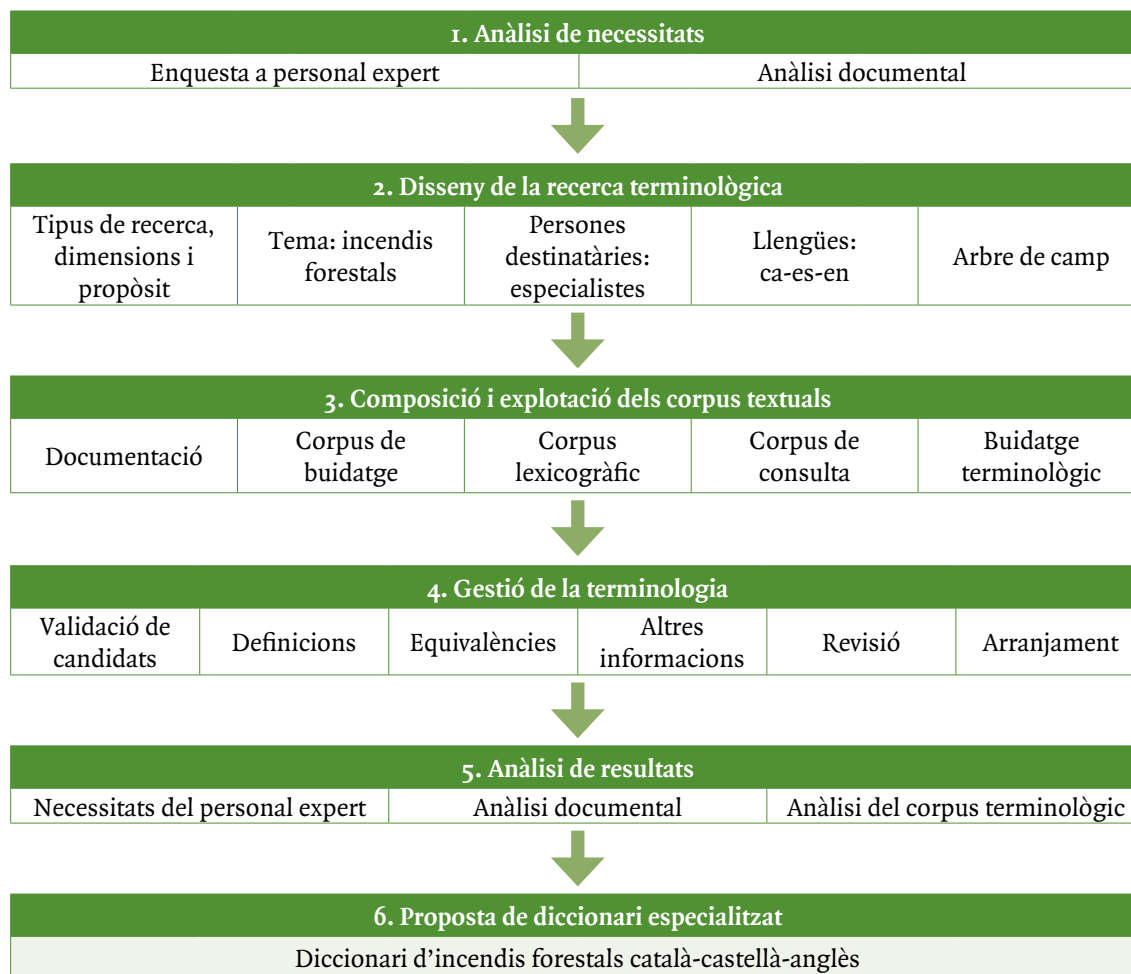


FIGURA 1. Seqüència metodològica de la investigació

- Propòsit: recopilar i descriure la terminologia dels incendis forestals, així com dissenyar una aplicació terminològica útil.
- Tema: incendis forestals.
- Persones destinatàries: personal tècnic especialista.
- Llengües: català com a llengua principal, i castellà i anglès com a llengües d'equivalència.
- Delimitació del camp: mitjançant l'elaboració d'un arbre de camp.

En tercer lloc, vam fer un intens treball de documentació sobre el camp per confeccionar els corpus textuais. En vam confeccionar tres, convenientment arranjats i codificats. El primer —de buidatge—, amb la finalitat específica d'extraure'n i analitzar-ne la terminologia, va estar compost de 859 textos especialitzats. Tots els textos eren escrits íntegrament en català i extrets de diferents contextos professionals vinculats als incendis forestals: juridicoadministratiu (lleis, decrets, ordres i resolucions específiques), tècnic (projectes, memòries, informes d'incendis, manuals i guies operatives), científicodivulgatiu (articles, butlletins especialitzats, actes de congressos i notes de premsa de bombers) i acadèmic (tesis doctorals i projectes de fi d'estudis). De manera complementària, vam confeccionar els altres dos, el corpus lexicogràfic i

el corpus de consulta per donar suport a la compleció del fitxer terminològic. Tant l'anàlisi del corpus com la gestió terminològica posterior, les vam dur a terme tecnològicament mitjançant l'estació de treball Terminus 2.0 del grup IULATERM.

En quart lloc, vam realitzar la selecció i validació dels candidats a termes, amb base a la nostra intuïció i als nostres coneixements especialitzats, seguint criteris semàntics, formals i quantitatius. A continuació, vam compondre el fitxer terminològic incorporant -hi els termes seleccionats i completant per a cadascun d'ells els camps següents: entrada, categoria gramatical, definició, context, equivalents, remissions, notes, fonts, freqüència absoluta, estructura gramatical i presència als diccionaris. Una vegada confegit el fitxer terminològic, el vam arranjament i supervisar en dues fases: revisió temàtica i revisió alfabètica.

En cinquè lloc, vam analitzar el corpus terminològic obtingut (1.134 termes) des dels punts de vista formal, semàntic, neològic i contrastiu, per tal d'establir-ne els trets singulars. En aquesta fase vam confrontar també les anàlisis realitzades a la primera fase, per esclarir les necessitats i les demandes del personal especialista en relació amb la terminologia.

En sisè i últim lloc, vam dissenyar una proposta de diccionari especialitzat per donar solucions concretes a les necessitats detectades. D'una banda, vam exposar una selecció de termes de l'especialitat, i de l'altra, vam suggerir l'estructura —tant externa com interna— que consideràvem més adequada per presentar-los.

3 Les necessitats terminològiques del personal expert en incendis forestals

Tota obra terminològica ha de tenir en compte el públic a qui s'adreça (Cabré, 1992). Partint d'aquesta premissa, el nostre estudi es va iniciar amb una exploració de les necessitats del personal expert en incendis forestals de l'àrea del domini lingüístic català prenent en consideració els dispositius d'extinció dels quatre pols territorials principals: Catalunya, País Valencià, Illes Balears i Andorra.

Volem fer ressaltar que conèixer les necessitats de les persones especialistes no es tracta d'una tasca senzilla, perquè aquestes són inevitablement subjectives, depenen de molts factors, no només de com són eixes persones i de les llengües que dominen, sinó també —i sobretot— del seu context professional i de les activitats concretes que duen a terme. Autors diverses assenyalen la importància de realitzar estudis empírics per poder esbrinar aquestes necessitats, així com els problemes reals amb què es troben (Bergenholtz i Agerbo, 2017; Fathi, 2015; Garcia-Soler i Rius, 2000; Koplenig, 2014).

3.1 Disseny de la recollida d'informació

Per poder recopilar les necessitats del personal especialista en incendis forestals de la nostra població objectiu, en relació amb l'ús de diccionaris, vam optar per utilitzar tres tipus d'instruments:

- una enquesta en línia;
- entrevistes a un petit grup d'especialistes;
- anàlisi de les obres terminològiques existents.

Ens interessava conèixer diverses dades del personal expert (sexe, edat i càrrec) i aspectes referents al seu entorn laboral (àmbit professional, àrea de treball, àmbit geogràfic i llengües de treball habituals), així com també saber si utilitzava —i com utilitzava— els diccionaris, tant generals com especialitzats: quines informacions considerava pertinents, les seves preferències en relació amb el format i l'accessibilitat, si les obres existents li resultaven útils i si trobava dificultats a l'hora de fer-les servir (vegeu la figura 2).

En el formulari vam utilitzar diversos tipus de respostes de manera estratègica: respostes tancades —tant d'alternativa simple (sexe) com d'alternativa múltiple (edat, àmbit, llengües, etc.)—, respostes d'opció múltiple (raons de consulta, format, informacions necessàries, etc.) i respostes obertes per recollir comentaris addicionals. Totes les preguntes van ser formulades tractant que foren precises i comprensibles. El conjunt de l'enquesta va ser elaborada en català estàndard, tot evitant dialectalismes i marques geogràfiques, perquè fora comprès i acceptat a tots els territoris.

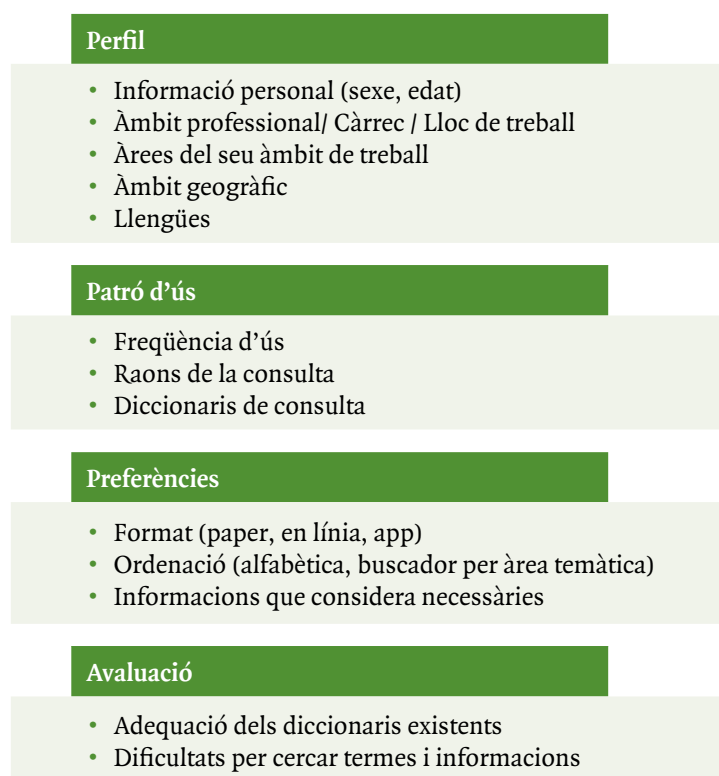


FIGURA 2. Disseny del qüestionari a personal expert

3.2 Persones participants

L'enquesta en línia fou contestada per vuitanta-dues persones, entre expertes i semiexpertes, representatives dels diferents territoris de parla catalana (Catalunya, País Valencià, Illes Balears i Andorra) i de diversos entorns professionals (tècnic, acadèmic, investigador i periodístic) relacionats amb els incendis forestals.

Pel que fa a les entrevistes, es van realitzar de manera oral a cinc especialistes de tres àmbits diferents (universitat, brigades i investigació). Les informacions obtingudes d'aquestes trobades van servir per complementar les informacions sobre les obres lexicogràfiques de referència en el sector, i també sobre com les usen, què hi busquen i quines mancances hi troben. Cal dir que la nostra experiència personal com a especialistes en incendis forestals és un dels pilars en els quals s'han fonamentat tota la cerca d'informació.

3.3 Anàlisi de dades del personal expert

L'enquesta i, d'una manera complementària, les entrevistes ens van proporcionar informació valuosa sobre el personal expert. A continuació, es presenta un compendi dels resultats, agrupats en quatre blocs:

I. Perfil

- **Sexe i edat.** Cal remarcar que la majoria de les persones enquestades van ser homes (75 %), pertanyents, fonamentalment, a dues franges d'edat: de 25 a 35 anys (48 %) i de 36 a 45 anys (39 %). Corroboem, doncs, la intuïció que teníem que l'àmbit professional relacionat amb els incendis forestals està format per personal jove i, a més, està fortament masculinitzat.
- **Àmbit professional.** Es distingeixen quatre sectors laborals diferenciats (vegeu la figura 3):
 - personal tècnic i directiu (68 %);
 - personal docent i investigador (17 %);
 - estudiantat/voluntariat (10 %);
 - periodistes del cos de Bombers (5 %).

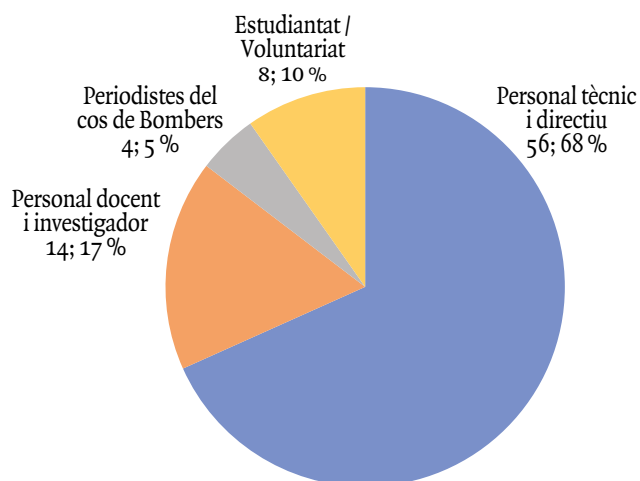


FIGURA 3. Àmbit professional del personal expert

Apreciem un clar predomini del personal tècnic i directiu (68 %), provinent, sobretot, dels serveis autonòmics de prevenció i extinció d'incendis de Catalunya i del País Valencià.

- **Subàrea temàtica.** Observem que totes les subàrees proposades dins del camp dels incendis forestals són productives i, en conjunt, força equilibrades (vegeu la figura 4). Una altra qüestió rellevant que advertim és que el personal expert habitualment es dedica a més d'una subàrea al mateix temps.

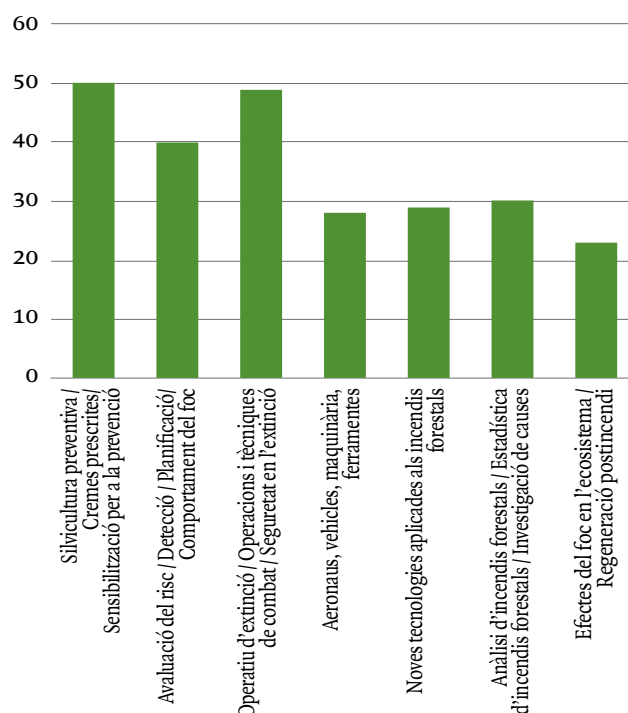


FIGURA 4. Subàrees temàtiques dels incendis forestals

- **Àmbit geogràfic.** La majoria de les persones enquestades pertanyen a l'àmbit geogràfic de les comunitats més grans, Catalunya (43 %) i el País Valencià (36 %); tenen menys representació les Illes Balears (10 %) i el Principat d'Andorra (1 %) (vegeu la figura 5).

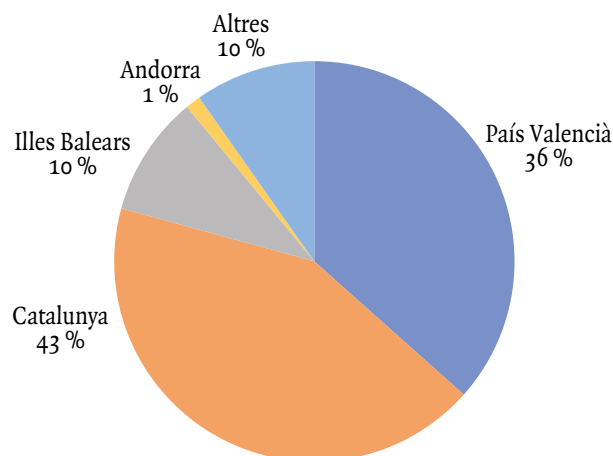


FIGURA 5. Àmbit geogràfic del personal expert

- **Llengües d'ús professional.** El personal expert manifesta que utilitza tres llengües de manera habitual en les comunicacions professionals (vegeu la figura 6), amb un lleuger domini del català (42 %), seguit del castellà (34 %) i, en menor grau, l'anglès (18 %). Tanmateix, usen també altres llengües com el francès, l'italià i l'alemany, amb uns percentatges d'ús significativament inferiors.

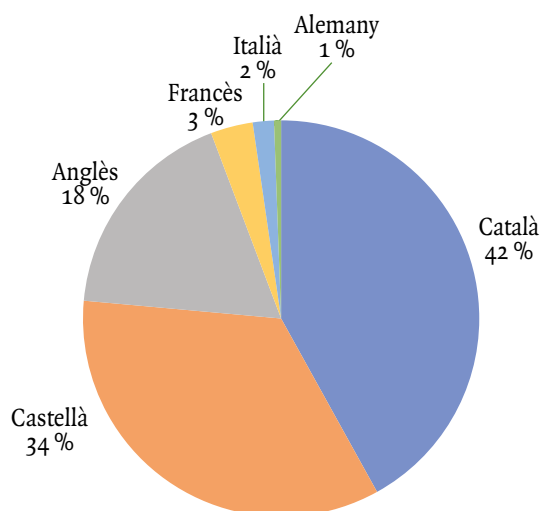


FIGURA 6. Llengües d'ús professional del personal expert

II. Patró d'ús

- **Freqüència d'ús.** Destaquem que el personal expert utilitza habitualment diccionaris (vegeu la figura 7): diàriament (15 %), d'una a dues vegades a la setmana (37 %) o entre una i dues vegades al mes (47 %).

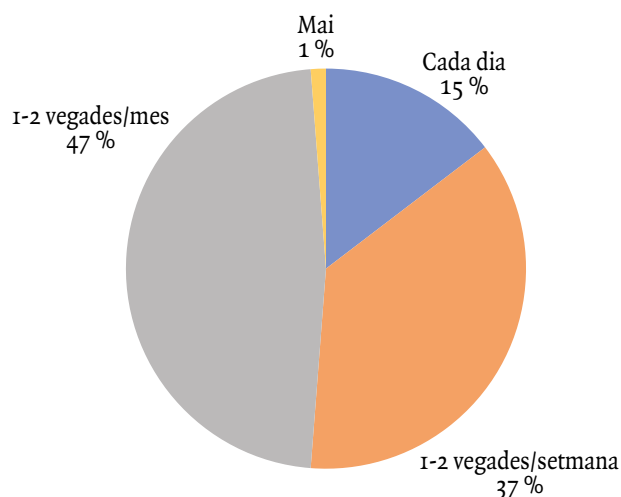


FIGURA 7. Freqüència d'ús dels diccionaris del personal expert

- **Format de diccionari.** El personal expert manifesta preferència pels diccionaris en línia (54 %), seguits de les aplicacions per a dispositiu mòbil (29 %). Tot i això, un 17 % continua utilitzant diccionaris impresos en paper.
- **Motivacions de la consulta.** Les activitats que motiven les consultes són molt diverses, però succeeixen fonamentalment escrivint en català (20 %) o en anglès (17 %), traduint (17 %) i llegint en anglès (17 %).
- **Tipus de diccionaris que consulten.** El personal expert fa cerques en tot tipus de diccionaris, però especialment en diccionaris generals bilingües o multilingües (26 %), en diccionaris generals monolingües (16 %), en diccionaris normatius (15 %) i en bancs de dades terminològics (13 %). En preguntar-los específicament pels diccionaris especialitzats sobre incendis forestals, una tercera part (32 %) palesa que no utilitza diccionaris especialitzats sobre incendis perquè, o bé no en coneix cap, o bé no els té a l'abast o bé no li resulten útils. Destaquem que el Cercaterm és la font especialitzada més consultada, seguida dels vocabularis forestals (Gil et al., 2010; Parés, 2010). En canvi, el Diccionari de bombers (TERMCAT. Centre de Terminologia, 2016) té menys acceptació.
- **Informacions que consulten.** Observem que el personal expert utilitza diccionaris fonamentalment per cercar la traducció d'un mot a una altra llengua (23 %), com s'escriu una paraula (22 %), què significa (20 %) i si és normativa (10 %).

III. Estructura idònia d'un diccionari sobre incendis forestals

- **Macroestructura.** Les persones especialistes expressen preferència per les versions en línia, presentades amb buscador (47 %) o de manera alfabètica (42 %). L'ordenació temàtica té, en canvi, molta menys acollida (10 %).
- **Microestructura.** Pel que fa a les informacions (vegeu la figura 8), el personal expert considera que els equivalents i les definicions són les més rellevants. El 78 % de les persones enquestades considera necessaris els equivalents a altres llengües i el 79 %, les definicions. Mostren, així mateix, un marcat interès pels sinònims (52 %), per les il·lustracions (48 %) i pels contextos d'ús (41 %) i, en menor grau (26 %), per les abreviatures i la subàrea temàtica. Despertem molta menys atenció (menys del 20 %) les categories gramaticals, l'etimologia, les marques de normalització, la procedència geogràfica o la neologicitat, els índexs i altres anotacions.

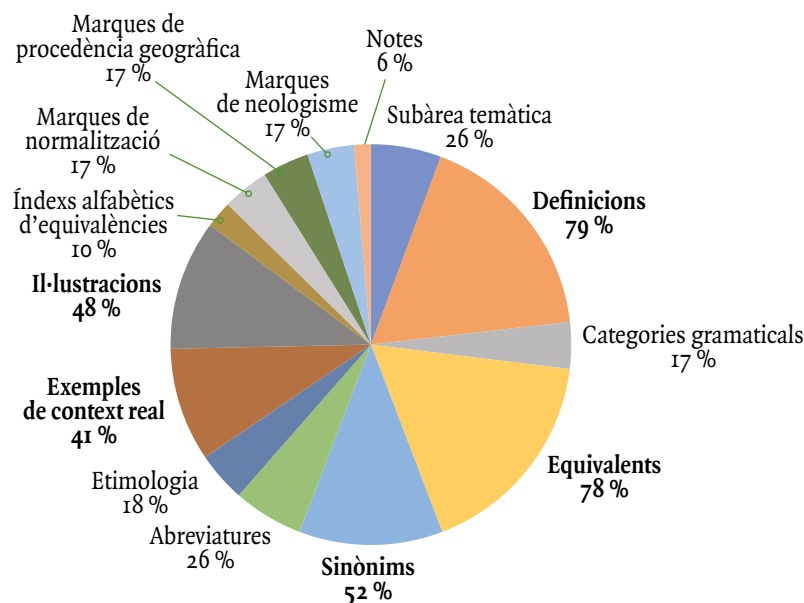


FIGURA 8. Informacions rellevants per al personal expert

Quant al tipus de terminologia recollida, s'interessen sobretot pels usos reals dels termes de l'especialitat (87 %), però també per l'estandardització d'aquesta terminologia en l'àmbit internacional (68 %) i, en un percentatge no gens menyspreable, per les recomanacions dels organismes normatius com ara l'Institut d'Estudis Catalans (37 %).

IV. Dificultats a l'hora de consultar les obres existents

El principal problema que relata el personal especialista és que amb una sola consulta no resolen la seva necessitat d'informació. Això ocorre generalment perquè han de consultar més d'un diccionari (el 44 % de les persones enquestades), però també perquè els termes consultats no hi apareixen (15 %), o bé perquè la informació que dona és excessivament bàsica, incompleta, inaplicable o obsoleta.

3.4 Anàlisi dels reculls terminològics existents

D'una manera complementària a l'anàlisi de necessitats del personal expert, vam estudiar les obres terminològiques existents —diccionaris, vocabularis, glossaris i petits reculls aplegats en documents tècnics— exclusives sobre incendis forestals. Després de valorar tots els reculls, vam bandejar els diccionaris forestals generalistes i altres diccionaris d'àrees afins com la meteorologia, les comunicacions o els bombers, perquè no s'ajustaven als objectius del tipus d'obres que volíem analitzar, per bé que es van incorporar al corpus de consulta. Així, també vam obviar alguns petits reculls i aquells d'autoria dubtosa.

Finalment, vam considerar un total d'onze obres referents de l'estandardització i la divulgació de la terminologia dels incendis forestals avalades per institucions referents en el sector, com el United States Department of Agriculture (USDA), el National Wildfire Coordinating Group (NWCG), el Global Fire Monitoring Center (GFMC), l'Australasian Fire and Emergency Service Authorities Council (AFAC) o el Centro para la Defensa contra el Fuego (CDF). A més a més, hi incorporarem tres projectes terminològics d'àmbit europeu, realitzats els anys 2009, 2012 i 2019.

A continuació, es presenten les onze obres terminològiques de referència analitzades (vegeu la taula 1).

L'estudi de les obres es va estructurar en tres blocs de paràmetres:

- Identificació: l'autoria, el propòsit, el lloc d'elaboració, el format d'edició i el grau d'actualització.
- Macroestructura: l'organització general del contingut i la facilitat d'ús.
- Microestructura: la quantitat i la qualitat de les informacions recollides.

D'entre les qüestions analitzades, destaquen les observacions següents:

- La procedència geogràfica de les obres és diversa, però majoritàriament anglosaxona (Estats Units, Canadà i Austràlia). En conseqüència, una gran part dels termes que contenen no són aplicables a la realitat professional de l'àmbit català.
- Totes les obres són adreçades a personal especialista.
- Totes les obres estan elaborades o avalades per institucions de referència en el sector d'àmbit internacional, d'entre les quals només n'hi ha una de procedència espanyola (CDF).

Obra	Any	Procedència	Codi	Llengües
Glossary of terms used in forest fire control (United States Department of Agriculture (USDA). Forest Service, 1956)	1939-1956	Estats Units	USDA	en
Wildland fire management terminology. Terminologie de la lutte contre les incendies de forêt. Terminología del control de incendios en tierras incultas (Food and Agriculture Organization of the United Nations (FAO) i Global Fire Monitoring Center (GFMC), 1999)	1986-1999	Internacional	FAO/GFMC	en, es, fr
Glossary of wildland fire terminology (National Wildfire Coordinating Group (NWCG), 2012)	1996-2012	Estats Units	NWCG	en
Firewords (Scott i Reinhardt, 2007)	2007	Estats Units	Scott	en
Glossary of forest fire management terms (Canadian Interagency Forest Fire Centre (CIFFC), 2021)	2002-2003-2021	Canadà	CIFFC	en, fr, es
Wildfire glossary (Australasian Fire and Emergency Service Authorities Council (AFAC). Rural and Land Management Group, 2009) / Busfire glossary (Australasian Fire and Emergency Service Authorities Council (AFAC). Rural and Land Management Group, 2012)	2009-2012	Austràlia	AFAC	en
Forest fire fighting terms handbook (Micillo, 2009)	2009	Europa	Micillo	en, es, fr, pt, it, el
European glossary for wildfires and forest fires (Stacey, 2012)	2012	Europa	Stacey	en
Forest fires multilingual glossary (Mediterranean Forest Fire Fighting Training Standardisation (MEFISTO), 2019)	2019	Europa	MEFISTO	en, es, fr, pt, it
International multi-lingual fire management terminology (Global Fire Monitoring Center (GFMC), 2010)	2010	Alemanya-Rússia	GFMC	en, ru, mn, de
Diccionario de incendios forestales: árabe - búlgaro - francés - inglés - portugués - rumano (Centro para la Defensa contra el Fuego (CDF), 2011)	2011	Espanya (Castella i Lleó)	CDF	en, es, fr, pt, ro, ar, bg

TAULA 1. Obres terminològiques exclusives de l'àmbit dels incendis forestals

- El propòsit majoritari de les obres és l'estandardització de la terminologia. L'única excepció és el CDF, elaborat amb la intenció de facilitar l'acollida lingüística de brigadistes no castellanoparlants.
- Només cinc obres disposen d'actualitzacions posteriors (USDA, FAO/GFMC, NWCG, CIFFC i AFAC). En relació amb les actualitzacions, cal dir que en la majoria dels casos no s'han fet de manera periòdica. Només dues obres han estat publicades en els últims deu anys (CIFFC i MEFISTO).
- El suport predominant és el de PDF en línia (NWCG, CIFFC, AFAC, Stacey, MEFISTO i GFMC), que se sol presentar amb les entrades ordenades alfabèticament.
- És difícil fer anàlisis quantitatives comparades, perquè només una de les obres publicades (FAO/GFMC) especifica el nombre d'entrades.
- Gairebé totes les obres recullen definicions i equivalents en altres llengües, només dues contenen sinònims (FAO/GFMC i Scott) i tan sols dues apleguen il·lustracions (Micillo i CDF).
- Tots els reculls contenen l'anglès i aproximadament la meitat contenen l'espanyol (FAO/GFMC, CIFFC, Micillo, MEFISTO i CDF), però cap no recull el català com a llengua d'equivalència.
A partir dels resultats, es pot afirmar que les obres analitzades no donen resposta a bona part de les necessitats manifestades pel nostre personal expert objectiu, ja que:
 - No contenen el català —ni altres llengües pròpies de l'Estat, com el gallec o l'euscar.
 - No recullen el castellà com a llengua pròpia ni hi donen prioritat en l'estructura del diccionari. El castellà apareix únicament com a llengua d'equivalència

i només en dos casos s'utilitza també en les definicions (CIFFC i MEFISTO). A més, atenent a la procedència de les obres, considerem que les traduccions castellanes són d'una fiabilitat qüestionable, amb l'única excepció de MEFISTO i CDF, les autories dels quals corresponen a persones especialistes de nacionalitat espanyola.

- No són aplicables al context espanyol. Si bé algunes obres derivades de projectes internacionals contenen l'espanyol com a llengua d'equivalència (MEFISTO), la terminologia que recullen no és totalment aplicable al nostre context perquè tant els ecosistemes naturals com els mateixos sistemes d'extinció (maquinària, tàctiques d'atac, dispositiu, jerarquia, etc.) són substancialment diferents.
- No són prou extenses. Malgrat les dificultats a l'hora de fer una anàlisi quantitativa de les obres, podem assenyalar que no responen a les necessitats en relació amb la quantitat de terminologia recopilada. De fet, una de les dificultats principals que relata el personal especialista és la dispersió de fonts, el fet de buscar en una obra i en una altra i no trobar-hi un terme. Com que no existeix cap obra de referència ni en català ni en castellà, el personal expert ha de consultar més d'una font i, tot i amb això, no té la certesa de trobar-lo.
- No estan prou actualitzades i, en alguns casos, contenen imprecisions o errors. Per al personal especialista, detectar absències importants o termes obsolets és símptoma de l'escassa qualitat de l'obra i un dels motius principals pels quals deixa de consultar-la.
- No contenen les informacions necessàries. Algunes obres no recullen definicions (Micillo i CDF) ni altres informacions —com categories gramati-

cals, il·lustracions o contextos d'ús— que de vegades requereixen.

Detectats els buits i les necessitats no satisfetes, vam considerar pertinent l'elaboració d'un diccionari d'incendis forestals català-castellà-anglès que recollís la terminologia del camp, de la manera més completa, precisa i adequada a la realitat del nostre territori. Per tal d'estructurar, i arrelar, la recerca d'aquesta terminologia, vam fer un treball intens de documentació.

4 La delimitació i l'estructuració del camp dels incendis forestals

Elaborar un diagrama conceptual dels incendis forestals requereix tenir coneixements previs sobre la matèria, però, sobretot, una visió panoràmica de tots els camps que abasta i de les relacions jeràrquiques que s'hi estableixen. Partint de la nostra experiència com a especialistes de la matèria, i d'un enfocament eminent tècnic, vam dissenyar un arbre de camp seguint una estructura basada en la tríada prevenció-extinció-postincendi, branques troncal que es corresponen amb els tres moments més importants d'intervenció sobre un foc forestal. Aquestes tres branques proporcionen l'esquelet bàsic del nostre diccionari i han servit per guiar l'arreplegada de termes, decidir quins n'havien de formar part —i quins no— i controlar la densitat terminològica de cada àrea. Cal dir que, malgrat la pretensió inicial de mantenir una certa proporcionalitat en la distribució de termes per àrees, de manera que no n'hi haguera cap d'infrarepresentada o de sobrerepresentada, la branca més abundant —amb diferència— ha estat la d'extinció i, d'entre les subbranques, la de comportament del foc (vegeu la taula 2).

Prevenció (24 %)	Extinció (59 %)	Postincendi (17 %)
Gestió agroforestal (14 %) aclarida successiva uniforme, pla de crema, foc prescrit, silvicultura preventiva, interfície urbanoforestal	Comportament del foc (27 %) carrera secundària, finestra meteorològica, longitud de flama, velocitat de propagació, cinturó tèrmic	Anàlisi d'incendis forestals (5 %) període de rotació del foc, zona homogènia de règim, investigació de causes, debriefing, cinquena generació d'incendis
Infraestructures i plans preventius (5 %) àrea tallafocs, via forestal, torre de guaita, punt d'aigua, mataespurnes	Operatiu d'extinció (18 %) director/directora d'extinció, brigada helitransportada, bomba forestal lleugera, torxa de degoteig, helicòpter Kamov	Efectes del foc en l'ecosistema (6 %) desertificació, erosió laminar, matèria orgànica pirogènica, microfauna, superhidrofobicitat
Gestió de risc d'incendi forestal (5 %) inflamabilitat, climatologia sinòptica, crema de rostolls, onada de calor, nivell de preemergència	Operacions de combat (14 %) atac paral·lel, lloc de comandament avançat, sufocar, desembarcar, foc tècnic	Regeneració postincendi (6 %) autosuccessió, restauració, espècie rebrotadora, vegetació climàtica, banc de llavors

TAULA 2. Camps semàntics de l'arbre de camp

Té lògica que l'extinció siga l'àrea més prolífica perquè des del punt de vista tècnic és precisament l'àrea temàtica més distintiva; semànticament, la més important i la més idiosincràtica, i també la que més quantitat de termes exclusius dels incendis forestals arreplega. La branca de prevenció és també bastant prototípica temàticament i una de les que més s'ha popularitzat entre el públic general gràcies a les campanyes de sensibilització. Quant a la branca postincendi, tot i que quantitativament és menys important, presenta una terminologia singular que convé no menystenir, la de l'ecologia del foc, i és la més representativa des del punt de vista científic.

Amb tot, s'ha de tenir present que dirimir l'àrea temàtica en alguns casos comporta una certa dificultat perquè algunes unitats són utilitzades amb propietat en dues o més subbranques i podrien ser considerades unitats interbranca. En aquests casos, s'ha adoptat el criteri d'assignar-les a la branca en què són usades

amb més freqüència o a la qual pertanyen més clarament des del punt de vista semàntic.

D'altra banda, hem delimitat algunes branques d'altres àmbits de la ciència forestal que estan força relacionats amb el camp dels incendis, com la nomenclatura botànica, la silvicultura o les repoblacions forestals. En aquests casos hem controlat molt l'entrada d'unitats perquè, malgrat que s'utilitzen amb una certa freqüència en el món dels incendis forestals, provenen d'àrees molt denses terminològicament i amb una entitat pròpia tan gran que podrien arribar a descompensar temàticament el recull.

A l'últim, volem subratllar que l'arbre de camp (vegeu la figura 9) ha tingut un paper important no només en l'elaboració del diccionari, sinó també en la fase de revisió, per controlar la pertinença dels termes al camp corresponent, per detectar absències conceptuals i, posteriorment, en el disseny de la proposta, per ampliar les possibilitats d'accés als termes.

- 1. PREVENCIÓ**
 - 1.1. Gestió agroforestal**
 - 1.1.1. Mosaic agroforestal
 - 1.1.2. Silvicultura preventiva
 - 1.1.3. Cremes prescrites
 - 1.2. Infraestructures i plans preventius**
 - 1.2.1. Punts d'aigua, xarxa viària
 - 1.2.2. Vigilància i detecció
 - 1.2.3. Mesures de defensa i autoprotecció
 - 1.3. Gestió del risc d'incendi forestal**
 - 1.3.1. Risc potencial: índexs i mapes
 - 1.3.2. Control d'ignicions
 - 1.3.3. Sensibilització i conscienciació social
- 2. EXTINCIÓ**
 - 2.1. Comportament del foc**
 - 2.1.1. Ignició, combustió, transferència de calor
 - 2.1.2. Meteorologia, topografia, combustibles
 - 2.1.3. Propagació del foc, incendis tipus, simuladors
 - 2.2. Operatiu d'extinció**
 - 2.2.1. Aeronaus, vehicles, eines, EPI
 - 2.2.2. Personal d'extinció
 - 2.2.3. Sistemes de comunicació
 - 2.3. Operacions de combat**
 - 2.3.1. Estratègia, tàctica, maniobra
 - 2.3.2. Comandament i emplaçament de mitjans
 - 2.3.3. Protocols de seguretat
- 3. POSTINCENDI**
 - 3.1. Anàlisi d'incendis forestals**
 - 3.1.1. Estadística d'incendis: nombre, superfície, causalitat
 - 3.1.2. Debrifing, paradoxa de l'extinció, generacions d'incendis
 - 3.2. Efectes del foc en l'ecosistema**
 - 3.2.1. Danys i processos erosius
 - 3.2.2. Ecologia del foc
 - 3.2.3. Hidrologia postincendi
 - 3.3. Regeneració postincendi**
 - 3.3.1. Resiliència i estratègies regeneratives
 - 3.3.2. Principis i tècniques de restauració

FIGURA 9. Arbre de camp

5 Conclusions

Hem estudiat el camp dels incendis forestals en els territoris de parla catalana, n'hem recopilat els textos i n'hem extret la terminologia des de la perspectiva del personal expert. Durant tot el procés hem posat molta atenció a les necessitats i expectatives dels professionals en relació amb la terminologia i hem tractat d'elaborar un producte terminològic útil, que faciliti la comunicació especialitzada.

Per concloure, volem destacar les observacions següents:

- La terminologia dels incendis forestals ha despertat un gran interès entre les institucions responsables de l'extinció d'incendis, especialment en l'àmbit anglosaxó. Observem en les obres existents un clar propòsit d'estandarditzar aquesta terminologia, principalment per a la comunicació internacional. En l'àmbit de l'Estat espanyol, malgrat que el personal expert manifesta interès tant pels usos reals de la terminologia com per la seva estandardització, no hi ha cap organisme que haja portat a terme aquest treball.
- Considerem que establir un estàndard terminològic augmentaria l'efectivitat de les comunicacions i l'operativitat, tant en l'àmbit d'un determinat dispositiu autonòmic com en l'arc catalanoparlant, i també en l'àmbit de tot l'Estat espanyol. En tot cas, com a pas previ a aquest treball prescriptiu, es feia necessari recopilar de manera sistemàtica tots els termes vius de la disciplina i descriure'ls convenientment. Des del nostre punt de vista, a partir d'aquí, sobre la base del plurilingüisme i de l'acceptació de la diversitat i la variació, es podria plantejar el treball d'harmonització de la terminologia entre les diferents institucions encarregades de l'extinció.
- En camps tan interdisciplinaris com el dels incendis forestals és cabdal parar molta atenció a la delimitació i l'estructuració del camp. D'aquesta manera es pot assegurar la pertinença dels termes seleccionats i garantir que seran útils per al personal expert, a més de comprovar que no es deixen buits conceptuals o denominatius significatius.

Des de la nostra perspectiva, la intervenció de persones especialistes és pertinent en l'elaboració d'un diccionari especialitzat, ja que la seva expertesa permet reforçar els principis de qualitat i adequació, tant en relació amb la qualitat de les informacions recollides com pel que fa a la utilitat del producte terminològic que es desenvolupa. Les necessitats i les preferències del personal especialista, així com les mancances que pugui detectar en les obres terminològiques, són informacions essencials que cal esbrinar a l'inici per dissenyar una obra idònia.

- Tanmateix, cal filar molt prim respecte al personal expert destinatari de l'obra, ja que no tot el personal especialista *per se* necessita les mateixes informacions. No només hi ha diferències significatives

entre persones expertes o semiexpertes, sinó que, a més, les necessitats tenen una relació directa amb l'entorn i amb l'activitat professional que es desenvolupa. En el nostre cas, el nostre perfil diana era fonamentalment el del personal tècnic dels serveis autonòmics d'extinció perquè es tractava del col·lectiu majoritari. Però fins i tot aquest perfil, mirat de prop, és heterogeni, ja que inclou tot un ventall de personal expert amb diferents nivells de formació que anirien des de la brigada fins a l'escala superior, d'inspecció, caps de serveis i analistes. Amb tot, vam decidir comptar igualment amb el professorat especialitzat en incendis, així com també amb el conjunt de personal investigador. En l'extrem de la consideració d'especialistes hi ha les persones aprenents d'especialista tant a la universitat com a la formació professional o a la formació en empresa privada que dona servei a les administracions autonòmiques. Considerem que l'elaboració d'un recurs específic per a algun d'aquests col·lectius requeriria una anàlisi de necessitats diferenciada. Finalment, malgrat que aquesta aplicació pugui ser útil per a professionals del periodisme i de la traducció vinculats a l'àmbit dels incendis forestals, remarcuem aquí que el recurs proposat no està dissenyat des de l'òptica de la divulgació i la traducció, sinó exclusivament des de la del personal especialista.

- El mètode de treball seguit ens ha permès identificar algunes qüestions essencials per al personal expert, que hem tingut en compte en el disseny del nostre diccionari:
 - Hem elaborat una obra major, extensa, de més d'un miler de termes, que abasta tots els subcampus de la disciplina recollits a l'arbre de camp.
 - El diccionari incorpora tres llengües d'ús habitual del personal especialista: el català, el castellà i l'anglès.
 - Hem recopilat termes vius i adequats a l'àmbit territorial estudiat perquè partíem d'un corpus de textos tècnics elaborats en el context espanyol i, més específicament, en el context catalanoparlant. Tots els textos han estat recopilats amb base a una sèrie de criteris coherents amb el perfil d'especialistes objectiu.
 - Hem dissenyat una microestructura fonamentada en les necessitats del personal expert, que conté números d'entrada, categories gramaticals, definicions, equivalents, sinònims i contextos d'ús. Així mateix, estem valorant la incorporació d'il·lustracions o d'infografies d'alguns conceptes.
 - Totes les informacions recopilades han estat revisades des del punt de vista dels especialistes per garantir-ne la precisió i l'actualitat.
 - En relació amb el format, estem valorant la manera més adequada de fer-lo accessible en línia amb un disseny que suporti diferents sistemes operatius i dispositius de consulta.

- Igualment, estem confeccionant la pàgina web del diccionari amb una macroestructura que permeti accessos múltiples a les entrades, mitjançant cerca avançada i ordre alfabètic, que siga comprensible i fàcilment operable per a les persones usuàries.
- Per acabar, hem considerat fer-ne actualitzacions regulars per mantenir-ne la vigència.

A manera d'epíleg, volem remarcar que els termes són un element fonamental de la ciència del foc. És a través dels termes que el personal expert és capaç d'es-

tablir les interconnexions cognitives necessàries per copsar tota la complexitat de la matèria. Es tracta d'un procés progressiu i sinèrgic: s'aprèn a gestionar el foc alhora que es van adquirint els termes que el defineixen. Tota persona especialista en incendis forestals n'ha de poder utilitzar la terminologia, d'una manera adequada i precisa, en totes les llengües del seu entorn professional. En aquest projecte, hem fet tot el possible perquè estigui disponible en una sola obra adaptada a les seves necessitats. 🌱

6 Bibliografia

- ANDRADE, Alfonso; LÓPEZ SANTALLA, Antonio; LÓPEZ GARCÍA, Marta (2019). *Los incendios forestales en España: Decenio 2006-2015* [en línia]. Madrid: Ministerio de Agricultura, Pesca y Alimentación (MAPA). <https://www.miteco.gob.es/es/biodiversidad/temas/incendios-forestales/incendios-decenio-2006-2015_tcm30-521617.pdf> [Consulta: març 2023].
- AUSTRALASIAN FIRE AND EMERGENCY SERVICE AUTHORITIES COUNCIL, RURAL AND LAND MANAGEMENT GROUP (2009). *Wildfire glossary*.
- (2012). *Bushfire glossary* [en línia]. <<https://www.afac.com.au/docs/default-source/doctrine/bushfire-terminology.pdf>> [Consulta: març 2023].
- BERGENHOLTZ, Henning; AGERBO, Heidi (2017). «Types of lexicographical information needs and their relevance for information science». *Journal of Information Science Theory and Practice* [en línia], 5 (3), p. 15-30. <<https://doi.org/10.1633/JISTaP.2017.5.3.2>> [Consulta: març 2023].
- CABRÉ, Maria Teresa (1992). *La terminología: La teoría, els mètodes, les aplicacions*. Barcelona: Empúries.
- (1999). *La terminología: Representación y comunicación*. Barcelona: Universitat Pompeu Fabra. Institut Universitari de Lingüística Aplicada.
- CANADIAN INTERAGENCY FOREST FIRE CENTRE (2021). *Canadian wildland fire management glossary* [en línia]. <<https://www.cifc.ca/publications/glossary>> [Consulta: març 2023].
- CASTELLNOU, Marc; RODRÍGUEZ, Lluís; MIRALLES, Marta (2003). «Replantant-se el futur de la prevenció i extinció d'incendis». *Jornades sobre Incendis Forestals i Recerca* [en línia] (Solsona, 12-13 juny). <https://interior.gencat.cat/web/.content/home/o30_arees_dactuacio/bombers/foc_forestal/jornades_recerca_cooperacio_internacional/articles_de_recerca_en_foc_forestal/articles_incendis_forestals/2003_Castellnou-et-al.-Solsona_REPLANTEJANT_extincio.pdf> [Consulta: març 2023].
- CENTRO PARA LA DEFENSA CONTRA EL FUEGO (2011). *Diccionario de incendios forestales: árabe - búlgaro - francés - inglés - portugués - rumano*. Lleó: Junta de Castilla y León.
- COSTA, Pau; CASTELLNOU, Marc; LARRAÑAGA, Asier; MIRALLES, Marta; KRAUS, Daniel (2011). *La prevenció dels grans incendis forestals adaptada a l'incendi tipus* [en línia]. Cerdanyola del Vallès: Generalitat de Catalunya. Departament d'Interior. Direcció General de Prevenció, Extinció d'Incendis i Salvaments. Unitat Tècnica del GRAF, Divisió de Grups Operatius Especials. <https://interior.gencat.cat/web/.content/home/o10_el_departament/publicacions/proteccio_civil/guia_la_prevencio_dels_grans_incendis_forestals_adaptada_a_l_incendi_tipus/docs/guia_la_prevencio_dels_grans_incendis_forestals_cat.pdf> [Consulta: març 2023].
- ESTOPÀ, Rosa (2013). *Les aplicacions terminològiques*. Barcelona: Universitat Oberta de Catalunya.
- FATHI, Besharat (2015). «Experts and Specialised Lexicography: perspectives and needs». *Terminàlia* [en línia], 9, p. 12-21. <<https://doi.org/10.2436/20.2503.01.61>> [Consulta: març 2023].

- FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS. GLOBAL FIRE MONITORING CENTER (1999). *Revision of the FAO wildland fire management terminology by the Global Fire Monitoring Centre (GFMC)* [en línia]. Food and Agriculture Organization of the United Nations (FAO). <<https://gfmcc.online/wp-content/uploads/GFMC-FAO-Fire-Management-Glossary-1999-edited-2018.pdf>> [Consulta: març 2023].
- GARCIA-SOLER, Jordi; RIUS, Lluís (2000). «Especialistes i diccionaris terminològics: una anàlisi de necessitats». *Terminologia e indústries da língua: Actas do VII Simpósio Ibero-Americano de Terminologia*, p. 14-17.
- GIL, Adriana (2022). «La terminologia catalana dels incendis forestals» [en línia]. Tesi doctoral. València: Universitat Politècnica de València. <<http://hdl.handle.net/10251/186183>> [Consulta: març 2023].
- GIL, Adriana; DELGADO, Rafa; GARCIA, Vicent; MACDONALD, Penny; LÓPEZ, Mercedes; BAYNAT, Maria Elena (2010). *Vocabulari forestal: català, castellà, anglès, francès* [en línia]. València: Xarxa Vives d'Universitats. Universitat Politècnica de València. <<http://www.upv.es/upl/U0552821.pdf>> [Consulta: març 2023].
- GLOBAL FIRE MONITORING CENTER (2010). *International multi-lingual fire management terminology (English, Russian, Mongolian, German)* [en línia]. <<http://gfmcc.online/literature/RUS-MON-GER-ENG-Glossary-Web.pdf>>.
- KOPLINIG, Alexander (2014). «Empirical research into dictionary use: A brief guide». A: *Using online dictionaries*, 2002 [en línia], p. 55-76. <<https://doi.org/10.1515/9783110341287.55>> [Consulta: març 2023].
- MEDITERRANEAN FOREST FIRE FIGHTING TRAINING STANDARDISATION (2019). *Forest fires multilingual glossary* [en línia]. <https://www.mefistoforestfires.eu/sites/default/files/annexes/forest_fire_multilingual_glossary_sp.pdf> [Consulta: març 2023].
- MICILLO, Gianfilippo (2009). *Forest fire fighting terms handbook* [en línia]. Corpo Forestale dello Stato. Dipartimento della Protezione Civile. Presidenza del Consiglio dei Ministri. <<https://gfmcc.online/wp-content/uploads/Forest-Fire-Terms-Handbook-EU-Fire-4.pdf>> [Consulta: març 2023].
- MORENO, José María; ROSA, D. de la; ZAZO, C. (2005). «Riesgo de incendios forestales». A: *Proyecto ECCE - Informe final: Evaluación preliminar de los impactos en España por efecto del cambio climático* [en línia]. Madrid: Ministerio de Medio Ambiente, p. 581-614. <https://www.miteco.gob.es/es/cambio-climatico/temas/impactos-vulnerabilidad-y-adaptacion/plan-nacional-adaptacion-cambio-climatico/evaluacion-preliminar-de-los-impactos-en-espana-del-cambio-climatico/eval_impactos.aspx> [Consulta: març 2023].
- MORENO, José María; URBIETA, Itziar R.; BEDIA, Joaquín; GUTIÉRREZ, José Manuel; VALLEJO, Vicente Ramon (2015). «Los incendios forestales en España ante el cambio climático». A: *Los bosques y la biodiversidad frente al cambio climático: impactos, vulnerabilidad y adaptación en España* [en línia]. Madrid: Ministerio de Agricultura, Alimentación y Medio Ambiente. <https://www.miteco.gob.es/es/cambio-climatico/temas/impactos-vulnerabilidad-y-adaptacion/cap34-losincendiosforestalesenespanaantealcambioclimatico_tcm30-70236.pdf> [Consulta: març 2023].
- NATIONAL WILDFIRE COORDINATING GROUP (2012). *Glossary of wildland fire terminology* [en línia]. <<http://www.nwccg.gov/pms/pubs/glossary/pms205.pdf>> [Consulta: març 2023].
- PARÉS, Eduard (2010). *Vocabulari forestal* (2a ed.). Barcelona: Institut d'Estudis Catalans. <<https://cit.iec.cat/VOCFOR/default.asp?opcio=1>> [Consulta: març 2023].
- PAUSAS, Juli G. (2012). *Incendios forestales*. Madrid: Catarata: Consejo Superior de Investigaciones Científicas (CSIC).
- SCOTT, Joe H.; REINHARDT, Elizabeth D. (2007). *Firewords: Glossary of fire science terminology* [en línia]. United States Department of Agriculture. <<http://www.firewords.net>> [Consulta: març 2023].
- STACEY, Robert (2012). *European glossary for wildfires and forest fires* [en línia]. European Forest Fire Networks (EUFOFINET). <<https://gfmcc.online/literature/EUFOFINET-Fire-Glossary.pdf>> [Consulta: març 2023].
- TERMCAT. CENTRE DE TERMINOLOGIA (2016). *Diccionari de bombers* [en línia]. <<https://www.termcat.cat/ca/diccionaris-en-linia/216>> [Consulta: març 2023].
- UNITED STATES DEPARTMENT OF AGRICULTURE. FOREST SERVICE (1956). *Glossary of terms used in forest fire control* [en línia]. <<https://naldc.nal.usda.gov/catalog/CAT87208605>> [Consulta: març 2023].

Diccionario gastronómico de hongos mexicanos: la relevancia de un diccionario gastronómico cultural y la variación terminológica

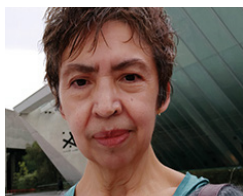
LAURA GABRIELA LINARES COLMENARES
Universitat Nacional Autònoma
de Mèxic (UNAM)
ORCID: 0000-0003-1155-1468
lauraglico@gmail.com

Laura Gabriela Linares Colmenares

Màster en Lingüística Aplicada per la UNAM. Autora del *Diccionario gastronómico de hongos mexicanos* (2021). Fundadora de la llibreria i casa d'estudis especialitzats en la cultura gastronòmica Conuco.



Marisela Colín Rodea. Doctora en lingüística aplicada per la Universitat Pompeu Fabra, s'ha especialitzat sobretot en l'ensenyament de llengües, els recursos lèxics i la lingüística aplicada com a àrea de coneixement de caràcter transdisciplinari. És professora de l'ENALLT-UNAM.



Amaranta Ramírez-Terrazo. Biòloga, màster en Ciències per la UNAM. Professora de la Facultat de Ciències de la UNAM. Estudiant de doctorat en Ciències Biològiques a la UNAM. La seva línia de recerca principal és el coneixement tradicional sobre els fongs silvestres.



AMARANTA RAMÍREZ-TERRAZO
Institut de Biologia, UNAM
ORCID: 0000-0001-9888-3224
a.ramirez.terrazo@gmail.com

Resumen

Este artículo presenta un estudio terminológico de la variedad de especies de hongos comestibles existentes en México. Partiendo de la teoría comunicativa de la terminología, de Cabré (1999), y planteando los nombres de los hongos como unidades terminológicas culturales, proponemos el primer Diccionario gastronómico de hongos mexicanos (DIGAHONMEX), el cual en su primera versión abarca hongos de la zona metropolitana, desarrolla una macroestructura y una microestructura dirigidas a cocineros y gastrónomos e incluye desde la información biológica básica hasta las formas adecuadas de cocción y recetas.

PALABRAS CLAVE: etnomicología; terminología; estudios de la alimentación; lexicografía; estudios culturales

Resum

Diccionario gastronómico de hongos mexicanos: la rellevància d'un diccionari gastronòmic cultural i la variació terminològica

Aquest article presenta un estudi terminològic de la varietat d'espècies de fongs comestibles existents a Mèxic. Partint de la teoria comunicativa de la terminologia, de Cabré (1999), i plantejant els noms dels fongs com a unitats terminològiques culturals, proposem el primer Diccionari gastronòmic de fongs mexicans (DIGAHONMEX), el qual en, la primera versió del qual abasta fongs de la zona metropolitana, desenvolupa una macroestructura i una microestructura adreçades a cuiners i gastrònoms i inclou des de la informació biològica bàsica fins a les formes adequades de coccio i receptes.

PARAULES CLAU: etnomicologia; terminologia; estudis d'alimentació; lexicografia; estudis culturals

Abstract

Diccionario gastronómico de hongos mexicanos: The importance of a culturally oriented gastronomical dictionary and the terminological variation involved

This article presents a terminological study of the variety of edible mushroom species existing in Mexico. Based on Cabré's Communicative Theory of Terminology (1999) and proposing the names of mushrooms as Cultural Terminological Units, we propose the first Gastronomic Dictionary of Mexican Mushrooms (DIGAHONMEX), which in its first version covers mushrooms from the metropolitan area. The dictionary developed a macrostructure and microstructure aimed at cooks and gourmets, and includes everything from basic biological information, proper cooking methods and recipes.

KEYWORDS: ethnomycology; terminology; food studies; lexicography; cultural studies

TERMINÀLIA 27 (2023): 20-31 · DOI: 10.2436/20.2503.01.186

Data de recepció: 06/02/2023. Data d'acceptació: 05/04/2023

ISSN: 2013-6692 (impresa); 2013-6706 (electrònica) · <http://terminalia.iec.cat>

1 Introducción

México es el segundo país en el mundo con mayor número de hongos usados como alimento (Garibay-Orijel, 2009; Ruan, 2007). Además, a la amplia diversidad de especies se le asocia una extensa variedad de nombres, algunos compilados en diccionarios de cocina mexicana. Sin embargo, se observaron diversos problemas en estas obras, como son: escasa información de cada entrada, información incoherente entre un artículo y otro, desorganización en las variaciones denominativas, información confusa sobre las maneras de cocinar y consumir los hongos, omisión de información relevante y específica para cocinar ciertas especies, entre otros. Por estas razones se toma la decisión de plantear un diccionario gastronómico especializado en hongos alimenticios mexicanos.

Para el diccionario se asumió la propuesta lexicográfica que toma los planteamientos expuestos en la teoría comunicativa de la terminología (TCT), de Cabré (1999), pero eso no lo excluye de ser una obra lexicográfica especializada, sino que supera este conflicto creando un diccionario que tiene como fin ser útil a un público específico que no tiene acceso a este tipo de obras. Por lo tanto, como afirma Edo-Marzá (2012): «la tradicional distinción entre lexicología y terminología y lexicografía y terminografía no resulta válida aquí por ser distinciones artificiales e innecesarias» (p. 111). En el caso del diccionario que presentamos aquí, más allá de caracterizar la obra como lexicográfica o terminológica se buscó definir el tipo de unidades: qué son las denominaciones de hongos en la gastronomía. A diferencia de otras áreas científicas, o incluso los propios nombres científicos de los hongos, que se han creado de manera específica para los estudios en la ciencia, los nombres de los hongos que se usan en la gastronomía tienen su origen en las comunidades donde los recogen y consumen. Estas denominaciones son parte del lenguaje general para esas comunidades rurales, pero han entrado en el lenguaje gastronómico por la relación que existe entre recolectores, comerciantes-recolectores y cocineros urbanos.

Debido a este fenómeno lingüístico y social que ocurre en la denominación de los hongos en el área gastronómica, se decidió también hacer un análisis de la variación conceptual y denominativa de acuerdo con: 1) la funcionalidad y 2) los factores sociales. El objetivo no fue buscar las causas de la variación, sino observar cómo existen correlaciones entre diversos factores que pueden incidir en tal variación. Para el análisis, que se realizó en Linares (2021), se tomaron cinco géneros: *Agaricus*, *Amanita*, *Ramaria*, *Russula* y *Morchella*, lo que representa el 10 % de los géneros totales recopilados en la investigación y un 38 % de las especies. En el presente artículo se exponen dos fenómenos del género *Agaricus*: 1) el caso del **san juanero** y 2) la influencia de la unidad terminológica cultural (UTC) *champiñón* y dos fenómenos del género *Morchella*: «¿Quién reco-

noce las **morillas**?» y «Los **olonanácatl** existían antes que las **morillas**».

El análisis de variación de los cinco géneros permitió conformar un corpus de **599 denominaciones** correspondientes a **124 especies**, entre las cuales se pudieron contabilizar **393 UTC** diferentes. De este corpus se realizó la obra prototípica sistematizada y con rigor lexicográfico que presentamos, de cincuenta entradas y sus variaciones denominativas: el *Diccionario gastronómico de hongos mexicanos* (DIGAHONMEX), volumen 1. La obra tiene una propuesta de macroestructura y de microestructura innovadora que responde a la necesidad de la comunidad gastronómica mexicana, al incluir las características biológicas básicas de las especies de los hongos, además de información cultural y gastronómica, como: usos frecuentes de cada hongo en las cocinas, niveles de cocción ideales, preparaciones recomendadas y recetas (las cuales se presentan de manera detallada en el artículo).

2 El Diccionario gastronómico de hongos mexicanos y las UTC

Elaborar un diccionario de hongos comestibles o alimenticios abre el panorama a la diversidad de tipos de diccionarios (culturales, enciclopédicos, científicos, etnobiológicos, etc.) y, ante estas posibilidades, el DIGAHONMEX se propone como un diccionario terminológico cultural.

Para el planteamiento del diccionario, aunque se denomina como una obra terminológica, resultó innecesario hacer distinciones estrictas entre si el diccionario era un diccionario de lexicografía especializada o uno terminológico, tal como plantea Alves da Conceição Costa (2015): «existen más similitudes que diferencias entre la Lexicografía y la Terminología, lo que prueba que son áreas complementarias y no excluyentes» (p. 276).

Si bien la razón por la que se denomina como terminológico se debe a que su teoría de base para la concepción, así como el análisis de las UTC, vienen de los postulados de la terminología —específicamente, de la TCT—, esto no excluye que el diccionario también se encuentre posicionado como una obra lexicográfica especializada. Por lo tanto, a pesar de otras diferencias, concordamos con lo que demuestra Alves da Conceição Costa (2015) en su investigación al afirmar que «los términos *diccionario terminológico* y *diccionario especializado* son sinónimos y que la variación denominacional se produce sin consecuencias cognitivas, dado que el uso de una u otra UT [unidad terminológica] no interfiere en la forma en que el consultante recibe información sobre ellas, ni en la comprensión que se obtiene del concepto transmitido» (p. 148).

La decisión de denominarlo *terminológico* está más relacionada con la TCT que se toma para la propuesta, así como con el uso de otras líneas de investigación

de la terminología, como la socioterminología, que aportaron claridad para la caracterización de los nombres de los hongos en la gastronomía como unidades terminológicas (UT). En este caso se consideraron los planteamientos de Faultish (1995), en los cuales afirma que las UT tienen un «contexto social determinante» y que muchas veces se encuentran en la frontera entre el lenguaje especializado y el general. No existe una línea rígida entre ambos; en el caso de las UT de los hongos gastronómicos, estas unidades tienen actores lingüísticos y extralingüísticos que las diferencian, desde su origen, porque, a diferencia de las UT usadas en la ciencia, no fueron creadas específicamente para una disciplina. En el caso de la micogastronomía, el contacto tan estrecho que existe con el trabajo del campo y con comerciantes rurales hizo que se fueran heredando denominaciones de comunidades de hongueros y se empezaron a emplear en la gastronomía urbana. Así, es imposible abordar de la misma manera un diccionario de UT que fueron creadas por científicos y para uso de científicos, que un diccionario de UT que se heredan de conocimientos vivos y son empleadas en áreas de especialidad relacionadas con dinámicas culturales.

Se proponen los nombres de los hongos como una UTC, ya que responden a un hecho cultural como menciona Lara (2016): «el significado de las palabras no es efecto de la imputación de un significante a una idea, un concepto, una cosa, sino que se forma junto a él, durante el proceso real de la adquisición de la lengua y en el crisol de la experiencia histórica y cultural» (p. 66).

En función de estas características de la UTC, también se buscó definir o redefinir qué era una obra terminológica, gastronómica y cultural. Se consideró como una obra donde se conservan sabores, olores y costumbres que han configurado la vida social del ser humano desde la cocina. Un diccionario gastronómico de una comunidad se convierte en un archivo social de la región, una memoria a corto y largo plazo a la cual el cocinero y gastrónomo puede ir para entenderla y, si desea, recrearla en sus platos. Este tipo de obras no requiere una información biológica tan detallada como en el caso de diccionarios académicos o científicos, aunque sí datos básicos que permitan comparar ciertas características taxonómicas, propiedades organolépticas e información sobre los métodos de cocción, maneras de consumo o propiedades de las especies para determinadas recetas.

Otro tema que resalta a la hora de elaborar un diccionario gastronómico cultural es que en la gastronomía existe una gran variedad de cocinas, por lo tanto, no solo fue necesario conceptualizar qué era un diccionario gastronómico cultural, sino cuál era el público del diccionario. Se establecieron tres categorías en el tipo de cocinas existentes: cocina rural, cocina doméstica y cocina profesional,¹ y a partir de allí se determinó que el DIGAHONMEX va dirigido a un público con conocimiento medio sobre la preparación de hongos, en cualquiera de estas categorías. Puede haber cocineros

y gastrónomos con alto nivel de instrucción que desconozcan la cocina micológica.

Tener claro en el tipo de público al que va dirigido un diccionario gastronómico cultural permite coherencia en la información incluida en cada entrada, y mejor selección de preparaciones y recetas que tengan grados de dificultad acordes al usuario.

También se debe recordar que las obras terminológicas, gastronómicas y culturales que se enfocan en alimentos silvestres —en este caso, los hongos— no son guías para reconocimientos de especies, pero sí una herramienta con información importante para cualquier chef o cocinero que no esté relacionado de manera profunda con el producto.

2.1 Partes del DIGAHONMEX

2.1.1 El corpus

El corpus está compuesto por cincuenta y seis textos: dieciocho son textos recolectados de cuestionarios, catorce de entrevistas; cuatro, de conversaciones semidirigidas, y veinte, de investigaciones etnomicológicas. En total se utilizaron dieciocho textos orales (32 %) y treinta y ocho textos escritos (68 %).

Se realizó la recolección de datos para crear este corpus y extraer las UTC de las especies que se utilizarían para la obra. Además, no existía un corpus del ámbito de la cocina ni había datos previos de registros de cocineros en la zona metropolitana que hablaran sobre hongos, razón por la que es fundacional en cuanto a este tipo de información en el campo lingüístico y terminológico en México.

Los registros orales se recogieron específicamente para la investigación, y para los registros escritos se usaron investigaciones etnomicológicas, ya que aportan información social y cultural pertinente para analizar la variación.

A su vez, de acuerdo con las subáreas de especialidad —que se concretaron en: recolección (R), recolección y comercio (RP), comercio (P) y cocina (C)—, se obtuvieron ocho textos de R (14 %), diez textos de RP (20 %), cinco textos de P (9 %) y treinta y dos textos de C (57 %).

Una vez recopilados todos los datos, se elaboró un catálogo del corpus compuesto de la siguiente manera: 1) tipo de registro, 2) código, 3) título del registro, 4) autor, 5) año, 6) selección de información y 7) tipo de texto.

2.1.2 Macroestructura

El DIGAHONMEX se confeccionó a partir de las características de libro impreso. Se consideró como un diccionario físico con la función de ser un vehículo para que los hongos tomaran su espacio en la cocina y en la mesa urbana contemporánea de manera tangible, haciendo honor a la gastronomía rural y ancestral.

A partir de estos principios, el diccionario presenta la siguiente macroestructura:

1. Advertencia: se da una breve explicación de los objetivos del diccionario y el riesgo que implica desviarse de estos. En ninguna circunstancia se puede usar como una guía para la recolección de hongos.
2. Los hongos en la gastronomía mexicana: un ensayo producto de la investigación realizada que describe la importancia de los hongos en la cultura y la gastronomía mexicanas, su historia y relación con los diccionarios y el impacto positivo, tanto social como cultural, que genera incluir hongos en el hecho alimentario.
3. ¿Qué es el DIGAHONMEX?: en este texto se hace una breve descripción del concepto de un diccionario terminológico en el área gastronómica y se explica la razón por la cual el DIGAHONMEX se caracteriza como este tipo de obra.
4. ¿Cómo se elaboró el diccionario?: se presenta un resumen de la metodología utilizada para realizar la investigación.
5. Manual de uso: se explica de manera detallada la información que compone la microestructura; se describen y presentan todos los campos del artículo para que los lectores puedan consultar el diccionario de manera eficaz. Una parte muy importante del manual es la ilustración de «Partes del hongo», ya que es indispensable tener este conocimiento para entender la información contenida en el diccionario.
6. Conjunto de entradas: del total de las especies que aparecieron en el corpus se hizo una selección con base en aquellas que tuvieron una aparición de, al menos, 5 % en los textos, y que por lo tanto representan las de mayor importancia cultural. El corpus total resultante es de cincuenta entradas. La información de entradas y variaciones denominativas se sistematizó en cuadros de acuerdo con el índice variación, y se descartaron las especies sin la mínima aparición.

La organización de las entradas es mixta. La primera división está hecha a partir de los géneros y, dentro de cada género, se organizan en orden alfabético las entradas correspondientes a cada especie.

Una vez seleccionadas las cincuenta especies, se escogió, por especie, la UTC con mayor ocurrencia para determinarla como la correspondiente a la entrada. En el caso de las UTC que eran las de mayor ocurrencia en dos o más especies, se seleccionó como entrada la especie en que la UTC tuviera el mayor número de apariciones en total y, para las otras especies, la UTC que seguía con el índice más alto en cada una de ellas.

La obra se consideró monolingüe, por lo tanto, las entradas corresponden a términos del español de

México y, de los vocablos que etimológicamente son de lenguas originarias, se han tomado solo los que se reconocen como préstamos.

7. Lista general de consulta de especies alimenticias: se agregó una lista de ciento veinticuatro especies, el total de las especies encontradas en la investigación, y sus variaciones denominativas.

8. Conoce al equipo DIGAHONMEX: en la parte final de la obra se presenta a cada uno de los miembros que participaron en la elaboración del diccionario.

2.1.3 Microestructura

De acuerdo con lo propuesto por Cabré (1999), las UT tienen características esenciales y adicionales. Las esenciales son intrínsecas a la definición del concepto de la UT, mientras que las adicionales se determinan de acuerdo con los intereses de la obra terminográfica o lexicográfica.

Así, considerando ambas características, el contenido de la microestructura del diccionario se ha dividido en tres componentes de contenido y la ilustración:

- a) Información lingüística: está integrada por la entrada, la marca de variación conceptual («VC») y dos campos: el nombre científico y la variación denominativa. Si los términos del campo de variación denominativa también tienen variación conceptual tendrán la marca «VCn». Al final de la página se indicará con «VC» la variación conceptual del término de la entrada y con «VCn» la variación conceptual de los términos de variación denominativa.
- b) Información biológica: describe las características macromorfológicas y de hábitat de cada una de las especies. Así mismo, la información biológica se refuerza con una ilustración elaborada por Julia Arvelaíz que permite apreciar detalladamente las especies y las características principales que las definen.
- c) Información culinaria: en esta sección se agrupan los datos del tipo de cocina donde se emplea el hongo, momento ideal para su consumo, olor, sabor y técnicas de preparación. Se debe recordar que existe subjetividad por parte de los investigadores en cuanto a las percepciones: tal es el caso de olor y sabor. Por tal razón es importante destacar que la información puede diferir de otros estudios.

Con el fin de lograr la estandarización se elaboró, con la colaboración de los especialistas en gastronomía Juan Escalona y Cristina Rubio, un sistema de códigos que permite tener parámetros claros del nivel de cocción de cada una de las especies. Debido a la importancia de la información sobre el consumo, cada entrada es de un color según si se puede comer crudo, con cocción ligera o con doble cocción (figura 1).

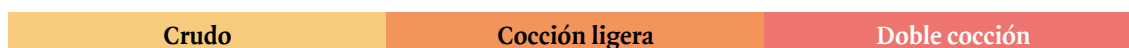


FIGURA 1. Códigos de nivel de cocción adecuada para cada una de las especies. Figura tomada de Linares-Colmenares (2021)

Así mismo, se estandarizaron siete técnicas principales aplicables a los hongos: asado, frito, sopa, tamal, quesadilla, guisado y conserva. De esta manera se puede brindar información estándar, y no tan extensa, en la que también se consideran preparaciones derivadas (tabla 1).

Por lo general, si en la definición aparece una sola técnica, como el caso de «frito», se sobreentiende que se le puede aplicar cualquiera de las tres preparaciones derivadas: capeado, tempura o *deep-fried*. En algunos casos, como el de «guisos», sí se especifica en el artículo la técnica con la variante de su preparación, ya que hay recomendaciones específicas para cada hongo. Algunos quedan bien en guisos como el mole, pero no con salsa verde, por lo tanto, es indispensable detallarlo. También, se agregó «guiso familiar» porque se reconoce que cada cocinero tiene una manera específica para cocinar de acuerdo con su contexto y/o herencia familiar o cultural.

Toda la información que contiene cada artículo se organizó en campos de la siguiente manera: entrada y color del hongo de acuerdo con el nivel de cocción, marca de variación conceptual (si es el caso), nombre científico, variación denominativa, definición e ilustración y especies de variación conceptual. En el caso de los nombres científicos, también se incluyeron marcas específicas: *aff.* (afín) se refiere a que puede ser la especie u otras que no están totalmente determinadas biológicamente, pero son morfológicamente muy similares, y *gpo.* (grupo), al conjunto de especies que tienen características biológicas y organolépticas similares y han sido agrupadas por métodos micológicos.

Para la definición se tomó en cuenta la metodología propuesta por Azevedo (2018), de modo que se dividió la definición en dimensiones y se estableció el orden de estas dimensiones de acuerdo con el interés del público al que va dirigido, como se presenta a continuación:

[descriptor común] + [uso cultural (gastronómico)] + [sabor] + [olor] + [consistencia] + [consumo] + [preparación] + [color y forma del píleo] + [textura y superficie del píleo] + [tipo y color del himenio] + [forma del estípite] + [textura del estípite] + [color del estípite] + [ornamento en el estípite] + [cambios de color] + [hábitos y hábitat de crecimiento]

Esta es la estructura genérica. En el caso de los géneros *Ramaria*, no tienen píleo sino ramificaciones, e igualmente ocurre en la variación con *Calvatia* y *Lycomperdon*, que se caracterizan por ser esporomas de forma esférica.

Por otro lado, cuidando el estilo y los recursos lingüísticos, se elaboró una plantilla estandarizada de redacción para cada una de las dimensiones. Dicha plantilla se flexibilizó de acuerdo con las características de los hongos o para los casos extraordinarios; sin embargo, se considera un recurso útil para evitar redacciones confusas. Esta plantilla de redacción se presenta a continuación:

Hongo (silvestre o cultivado) comestible en su estado (joven y/o maduro). Empleado en la cocina (doméstica / rural), (así como en la restauración). La consistencia es (). Tiene sabor () y olor de (). Se puede consumir () y se recomienda prepararlo (). El píleo es (color) y (textura) tiene una forma () con el centro (), borde (). El himenio (forma / textura y color). El estípite es (forma, textura y color). Crece en bosques de () y de forma ().

Por último, cada entrada se acompaña de una ilustración, ya que, como parte de la obra lexicográfica, las imágenes tienen el objetivo de ser un complemento de las definiciones. En este caso la imagen cobra importancia debido a que el aspecto físico del hongo es importante para el público al que va dirigido el diccionario. La ilustradora, en su proceso creativo, destacó las características distintivas de cada hongo para relacionar de manera muy precisa su aspecto físico con la definición. Para la ilustración, la artista Julia Arvelaiz empleó la técnica de la acuarela. Cada imagen está inspirada en las fotos de cada especie y en la literatura micológica especializada. Las imágenes son únicas y elaboradas de manera exclusiva para el diccionario.

Así, con esta información se confeccionó la microestructura y se diseñaron los artículos de cada entrada. En la figura 2 se presenta un ejemplo de los artículos terminológicos del DIGAHONMEX, en donde se aprecia de manera clara la organización del contenido.

Preparación principal		Variantes		
asado	a la leña	a la plancha	al gril	
frito	tempura	capeado	<i>deep-fried</i>	
sopa	caldo	a la mexicana	con pasta	
tamal				
quesadilla				
guisado	mole	salsa verde	salsa roja	guiso familiar
conserva	fermentado	encurtido		

TABLA 1. Diversidad de técnicas aplicables en la cocción de hongos. Tabla tomada de Linares-Colmenares (2021)

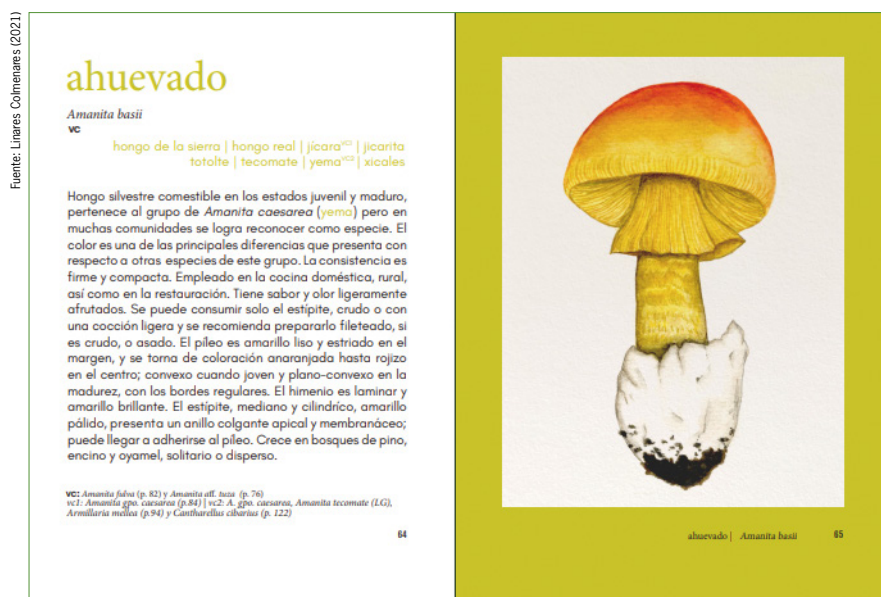


FIGURA 2. Modelo de los artículos terminológicos en el DIGAHONMEX

3 La sistematización de la variación denominativa y de la variación conceptual en un diccionario gastronómico cultural

Uno de los grandes desafíos del diccionario terminológico cultural es organizar de manera clara las variaciones terminológicas, considerando que las UTC vienen de contextos dinámicos y suelen tener una alta variación. Para el diccionario se consideraron las variaciones conceptuales y denominativas tomando como referencia las definiciones propuestas por Freixa (2002), que expone que las variaciones denominativas son «el fenomen pel qual a una mateixa noció li corresponen diverses denominacions» (p. 54) y que la variación conceptual «significa per a nosaltres variació en un mateix concepte: creiem que en el procés de denominació, un mateix concepte pot ser abordat de maneres diferents i que per aquesta raó poden aparèixer denominacions diferents. I, com hem dit, aquestes denominacions per a un mateix concepte tenen la virtut de permetre observar la part (o les parts) del contingut conceptual que ha pres més rellevància en un context determinat». (p. 55)

Para lograr una sistematización en cuanto a la variación conceptual, se utilizó la marca «VC» para indicar si la entrada tenía variación conceptual, y se ubicó en un campo debajo de la entrada (figura 3).

En el caso de la variación denominativa, se incluyó un campo donde van compiladas todas las variaciones denominativas de la especie correspondiente a la entrada (figura 4).

En el caso de las variaciones denominativas también se marcaron los términos que tenían a su vez variación conceptual. Para indicar este fenómeno se han señalado con la marca «VC» en fuente regular, en mayúsculas, en la parte superior de los términos, acompañada de un número (n). Así, la marca «VCn» indica que esos términos son una variación denominativa de la especie correspondiente a la entrada, pero que también tienen una variación conceptual (figura 5).

Al final del artículo lexicográfico se encontrará una sección donde aparecen las marcas «VC» y «VCn» acompañadas de los nombres científicos de las otras especies que también son denominadas con el término de la entrada, y de las especies que corresponden, igualmente, a los términos de la variación denominativa (figura 6).



FIGURA 3. Modelo de marcación de variación conceptual



Variaciones denominativas de **san juanero**

FIGURA 4. Modelo del orden de la variación denominativa de cada entrada



Marca de variación conceptual de los términos de variación denominativa para la especie

FIGURA 5. Modelo de la marcación de variación conceptual de la entrada y de los términos de la variación denominativa

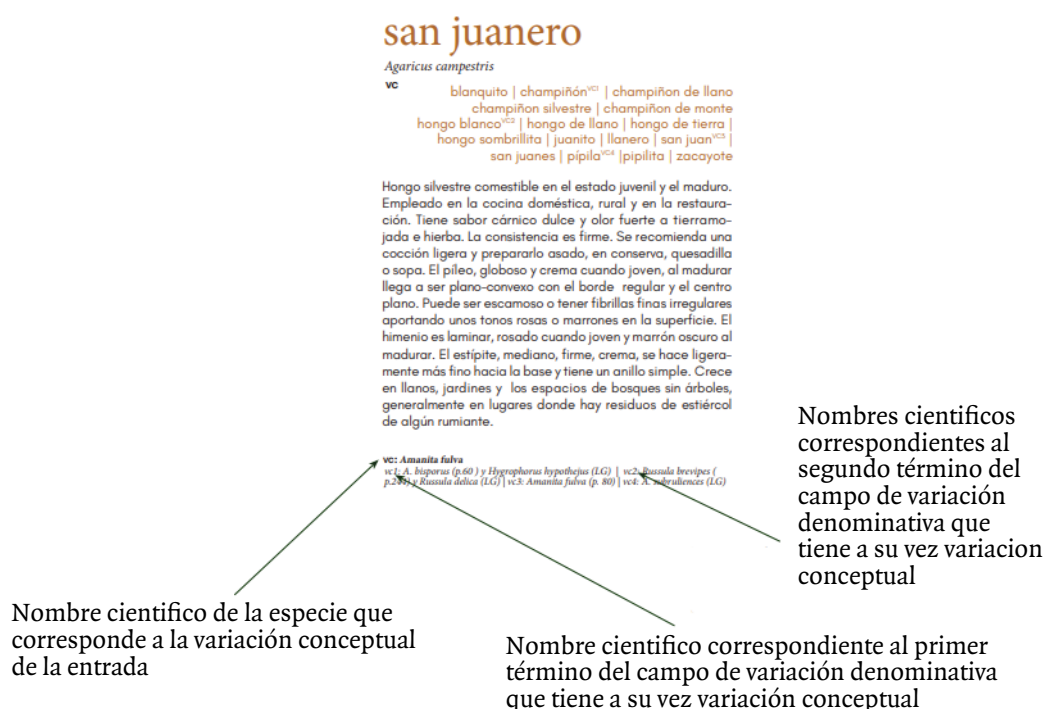


FIGURA 6. Modelo de la sección de los términos científicos correspondientes a la variación conceptual de la entrada y a las variaciones conceptuales de los términos de las variaciones denominativas

4 La importancia de la variedad terminológica en el diccionario gastronómico cultural: un análisis socioterminológico

Incluir la variación terminológica implica tener que poseer mucho más que un mero conocimiento lingüístico, por ello, previo a lograr la organización de las variaciones en la microestructura, se hizo un análisis (Linares, 2021) a partir de las líneas de investigación de la socioterminología y la propuesta de variación funcional de Cabré (1999).

Para la primera parte del análisis, la funcional, se hizo una división en las subáreas de la gastronomía que ya se han descrito anteriormente: cocina (C), recolección (R), comercio (P) y recolección y comercio (RP). Una vez hecha esta propuesta, la información del corpus se sistematizó según cómo eran nombrados los hongos en dichas subáreas.

Por otro lado, se realizaron fichas socioterminológicas con los siguientes datos: año del texto (en este caso, el texto puede ser tanto oral como escrito), segunda lengua, zona geográfica, contacto con lenguas originarias, contacto con personas de otras subáreas y estudios realizados.

Al tener sistematizada la información del corpus general de la investigación, de las tablas de las subáreas y de las fichas socioterminológicas, se seleccionaron cinco géneros de hongos: *Agaricus*, *Amanita*, *Ramaria*, *Russula* y *Morchella* (lo que representa el 10 % de los géneros totales de la investigación y un 38 % de las especies) y se hizo un análisis descriptivo y social con los datos de las fichas socioterminológicas y los datos de las tablas del análisis funcional. Es importante aclarar que el objetivo del análisis no fue buscar las causas de la variación —debido a que este tema ya ha sido ampliamente tratado por Freixa (2013)—, sino observar cómo existen correlaciones entre diversos factores y cómo estos pueden incidir en la variación terminológica. Por lo tanto, se fueron caracterizando diversos fenómenos, de los cuales presentamos algunos ejemplos de los géneros *Agaricus* y *Morchella*.

4.1 Análisis de la denominación en el género *Agaricus*: el caso del san juanero y la influencia de la UTC champiñón

En este género se agruparon cuatro especies y *Agaricus bisporus* con dos variedades. De las siete, cuatro se registraron con variación denominativa y tres no, ya que solo tuvieron una aparición en los textos. Tomando las cuatro especies que presentaban variación, se observaron diversos fenómenos lingüísticos que se clasificaron de la siguiente manera: 1. *A. bisporus* - **champiñón**, 2. El **san juanero** y la influencia de la UTC **champiñón** y 3. **Portobello**, **champiñón** y **cremini** ¿la misma especie?

El fenómeno denominado «el **san juanero** y la influencia de la UTC **champiñón**» involucra dos especies: *A. campestris* y *A. bisporus*, este último conocido generalmen-

te como **champiñón**, con características organolépticas muy similares al *A. campestris*, pero su principal diferencia reside en que uno es silvestre y el otro, cultivado.

Durante la recolección de datos se obtuvo como resultado que en la subárea cocina era mucho más conocido y usado *A. bisporus*, con veinticinco apariciones, frente al *A. campestris*, con tan solo tres. Lo interesante es que el *A. bisporus* no es un hongo que tenga mejores propiedades que el *A. campestris* para cocinar, sino que está relacionado con su disponibilidad en el mercado e industrialización, que ha hecho que cocineros y chefs —contemplando que gran parte de los informantes del corpus de cocina provienen de la cocina profesional— conozcan el de más fácil obtención y opten por él. Es cierto que el *A. campestris*, al ser silvestre, solo se puede usar durante la temporada de lluvias, pero no debería influir en el conocimiento de la especie en sí, aunque solo se cocine en una época del año.

Además de esta diferencia de aparición, se observa el fenómeno sobre la estandarización de los nombres. Los hongos que forman parte de la despensa de la cocina urbana o profesional y están industrializados presentan mayor estandarización. Tal es el caso del *A. bisporus*, el más usado en la cocina dentro del género *Agaricus*. A pesar de su alta aparición, solo se registraron tres UTC diferentes en todas las subáreas: **champiñón**, **regular** y **parís**. Es importante destacar que muchos de los comerciantes y recolectores de los hongos silvestres también venden hongos cultivados, pero, como se puede concluir por los datos recogidos, han conservado el término que se les ha asignado en la industria. A diferencia de la pequeña variación denominativa de *A. bisporus*, el *A. campestris*, con solo dieciséis apariciones como especie, tuvo un registro de diecisiete UTC diferentes: **blanquito**, **champiñón**, **champiñón silvestre**, **champiñón de llano**, **champiñón de monte**, **hongo blanco**, **hongo de llano**, **hongo de tierra**, **hongo sombrillita**, **juanito**, **llanero**, **san juan**, **san juanes**, **san juanero**, **pípila**, **pipilita** y **zacayote**. De estas UTC que se obtuvieron, se resalta que aparece también el término **champiñón**, y al contrario de lo que se podría creer de que estos usos corresponden a la subárea cocina, las UTC como **champiñón**, **champiñón silvestre**, **champiñón de llano** y **champiñón de monte** aparecieron en comercio y recolección y en comercio. La única que apareció en cocina fue **champiñón silvestre**.

Por lo tanto, se hizo un acercamiento a los datos de las comunidades de recolección donde se habían obtenido estas UTC y se encontró que en algunos de los textos donde está registrado **champiñón** o donde esta UTC forma parte de una UTC compuesta, se presentan también los usos **hongo blanco** y **hongo de llano**, de tal manera que no se utiliza exclusivamente el término **champiñón**, sino que este forma parte de la variación denominativa, de manera que coexisten **champiñón**, **hongo blanco** y **hongo de llano** como términos en una misma comunidad.

El uso de la UTC **champiñón** parece tener relación con la frecuente comercialización del hongo cultivado *A. bisporus*, que es el denominado champiñón originalmente. Aunque las comunidades que usan estas UTC no reportan comercio directo con personas de la urbe, sí se encuentran aledañas a otras comunidades que venden sus hongos para personas externas a su territorio. Por lo tanto, el uso del término champiñón para el *A. campestris* se extiende en la zona, ya que los comerciantes lo usan para entenderse con los compradores, y así ha empezado a formar parte de los términos utilizados por recolectores y comerciantes de las zonas, sin sustituir a las UTC propias de las comunidades originarias.

Tanto el uso de **hongo blanco** y **hongo de llano** parecen ser traducciones literales de las denominaciones en otomí y mazahua que forman parte de las lenguas de estas comunidades, en otomí-pame: *r taxi ju* ('hongo blanco') y *k' e charak'u* ('hongo de llano').

Resulta importante evidenciar que cuando un hongo proviene del lado occidental prevalece como UTC única la asignada desde la industria, que se estandariza; tal es el caso de uso de la UTC champiñón para el *A. bisporus*, aun en las zonas de recolección y comercio. Al contrario que cuando se introduce un hongo de comunidad originaria como el *A. campestris* a la cocina urbana: en tal caso, aunque se adopte el uso de alguna UTC de los recolectores y comerciantes, se sigue dando mayor uso en la urbe a la que ha sido propuesta en la cosmovisión occidental.

Uno de los casos particulares en donde se evidencia la interacción entre lo occidental y lo originario es la UTC **champiñón de llano**. Destaca porque, aunque esta construcción pareciera muy genérica del español, integra una unidad léxica de influencia del español peninsular (champiñón) y otra que es usada en el español de América o del español de México (llano), y que en muchas ocasiones es traducción del término originario en diferentes lenguas.

Para analizar este caso se tomaron como referencias las UTC que compila *Etnomicrología en Castilla-La Mancha*, una investigación etnomicroológica que lleva más de veinte años recabando datos en el sector, así como portales web españoles como *Guía de Setas y Hongos de Navarra*, *Fichas micológicas de la Cofradía Gastronómica Vasca* y la página comercial *Cestas y setas*. En esta búsqueda se encontró que el *A. campestris* se denomina **champiñón silvestre** o **champiñón de campo**, mientras que no se documentó la UTC **champiñón de llano**. Incluso cuando se busca llano en el *Diccionario de la lengua española* no es hasta la acepción dieciséis que se encuentra solo una hipervinculación a llanura y es en esta entrada donde se define el concepto de lo que se entiende como llano en México.

Lo que se observa es que llano no es un vocablo común en el español peninsular, al menos, en las fuentes consultadas. Por lo tanto, en el caso de los hongos es muy curioso este fenómeno: surgen UTC que combinan tantas palabras del español peninsular con una del español de México. Este fenómeno también se da en otras especies con un uso muy extendido en el ámbito urbano o en la cocina occidentalizada, tal como ocurre con el **quismón de huevo**: quismón, una palabra que ingresa al español de México del otomí y se une con huevo, una palabra común para la denominación de hongos en el español peninsular.

Con respecto a la UTC que tiene registro en la subárea cocina, se observa cómo **san juan** y los demás derivados de este término —**hongo san juanero**, **hongo de San Juan**, **juanito**— se relacionan más con el campo y se mantienen por su importancia cultural, más allá del sector profesional. Es curioso que el término con mayor ocurrencia para *A. campestris* precisamente sea **san juan**, un neologismo que se origina en los sectores rurales y está asociado a fiestas religiosas en la época de lluvias, cosa que también indica que el *A. campestris* tiene mayor importancia en la cocina rural.

En la época contemporánea algunos de los hongos se denominan con términos relacionados con el catolicismo, pero investigaciones etnomicroológicas, como la de Moreno Fuentes (2013), e investigaciones alimentarias, como las de Guía-Zaragoza (2016) y Bak-Geller (2013), reconocen que en el desarrollo de lo que sería la «cocina mexicana» existieron alimentos que fueron expulsados de los fogones porque no eran parte de la cosmovisión católica, entre ellos, los hongos. Sin embargo, su consumo se mantuvo de manera muy discreta y terminaron por adoptar nombres relacionados con el período de la colonización.

El nombre de *san juan*, *san juanero*, etc. se asocia al *A. campestris* por la fiesta de San Juan, pero incluso esta fiesta está íntimamente relacionada con las tradiciones prehispánicas como la fiesta de Tlaloc (Broda, 2019). El Sistema de Información Cultural de México establece que la fecha oficial de la fiesta de San Juan Bautista es el 24 de junio. Sin embargo, cada población establece, de acuerdo con las otras, unos días para celebrarla. En todos los casos se trata de fechas cercanas, las cuales coinciden, según Broda (2019), con el inicio de las fiestas de Huey Tozoztli, que se celebran en el cuarto mes del calendario solar mexicano y que dan inicio a las fiestas de peticiones de lluvias y otras ceremonias relacionadas con los ciclos agrarios, hasta llegar a las propias fiestas de Tlaloc. Este tipo de ceremonias y rituales se realizaban en el Valle de México, que corresponde al sector tomado para la investigación.

4.2 Análisis de la variación denominativa en el género *Morchella*: los casos «¿Quién reconoce las morillas?» y «Los olonanácatl existían antes que las morillas»

4.2.1 Los olonanácatl existían antes que las morillas

En el género *Morchella* se encuentran seis especies, de las cuales dos, *M. costata* y *M. deliciosa*, no presentan variación y tuvieron solo una aparición en el corpus; otra, *M. rufobronea*, solo tuvo una variación individual, y las otras cuatro especies, *M. elata*, *M. esculenta*, *M. angusticeps* y *Morchella* sp.1/ sp.2, sí presentaron variación denominativa.

La especie que tuvo mayor aparición en el corpus fue *M. esculenta*, la cual se encontró en algunos textos registrada científicamente como *M. cónica*, aunque se refiere a la misma especie. El porcentaje de variación denominativa en esta especie tan solo fue del 32 %; resultados similares se dan en las especies de los otros géneros analizados, en los cuales las especies con mayor uso presentan tendencia a estandarizarse. Se debe recordar que no es un factor determinante, pero sí se revela en el corpus una correlación entre una mayor aparición de las especies. Sin embargo, las características del género y la subárea donde se use la especie también influyen en el fenómeno.

La UTC **morilla** fue la que tuvo mayor ocurrencia en *M. esculenta*, con catorce apariciones, doce de las cuales fueron recogidas en la subárea cocina. Es importante destacar que de las doce encontradas en cocina nueve corresponden a textos de la prueba piloto, en donde la imagen de esta especie se aportaba como estímulo para que los encuestados respondieran qué UTC usaban para dicha especie. Este detalle es importante, porque si se observan las *Morchella* spp. aparece también **morilla** como la UTC con mayor registro. Lo que se evidenció durante las entrevistas es que la mayoría de los cocineros no distinguen las diversas especies del género *Morchella* y las denominan con el etnotaxón **morilla**, lo que lleva a suponer que, en el caso de la prueba piloto, si se les hubiese mostrado alguna foto de una *M. elata* u otra especie de *Morchella*, también la habrían llamado **morilla**.

El término **morilla** refleja la cocina occidental, tal como se muestra en las Fichas micológicas de la Cofradía Vasca de Gastronomía, en la Guía de hongos y setas de Navarra y en la página Cestas y setas. Incluso el uso de la UTC **morilla** también se relaciona con otras lenguas, como el inglés **morels**, tal como lo reporta EncicloVida.

En la denominación general de todas las especies del género *Morchella*, **morilla** tuvo aparición en veintinueve textos de cocina frente a cinco en las otras subáreas. En el caso de la subárea de recolección y comercio, en todos los registros que se tiene su ocurrencia (el término **morilla**) todos los usos se dan como variación de otras UTC relacionadas cognitivamente con referentes del campo como el **maíz** y la **panza** (en este caso refie-

re a la panza de res, que también es comestible). No se registró ningún texto en recolección y comercio, recolección o comercio en las que solo se usara **morilla** de manera estandarizada. Eso es porque parte de la variación denominativa que incluye **morilla** se origina por la influencia extranjera de la cocina de hongos, tal como se patentiza en algunos registros históricos como el de Melchor Ocampo, en su texto *Cocina francesa* o en los apuntes del *Nuevo cocinero mexicano*.

En el caso de la subárea de recolección se encontró la especie *M. esculenta* en cinco textos de recolección y dos textos de comercio. En estas subáreas, el término que tuvo mayor ocurrencia para la especie fue **pancita**. En otro texto de recolección y comercio se tiene el uso de **morilla**, pero con la variación denominativa en náhuatl: **olonanacatl**, **olo** ('mazorca') y **nanacatl** ('hongo').

Este texto de la subárea recolección y comercio donde se registraron los dos usos **morilla** y **olonanacatl** proviene de una comunidad donde ciertos pobladores hablan su lengua originaria náhuatl y también español. Por lo tanto, se observa que, de las veinticinco especies de hongos que se usan en la comunidad, tres se registran como UTC que son préstamos del náhuatl (**olonanacatl**, **tlazonanacatl** y **xoclaclich**). El uso de estos términos en la comunidad para los otros dos hongos no tiene variación; solo en el caso de la *M. esculenta* tiene las UTC **olonanacatl** y **morilla**.

El uso de los préstamos del náhuatl en este texto puede estar relacionado con que ciertos hongos tienen un alto valor cultural y por eso se mantienen los términos originarios; pero, a su vez, la amplia demanda, como en el caso de *M. esculenta*, que representa el segundo producto con mayor valor económico de la comunidad de donde se toma el texto, de acuerdo con Arteaga Martínez et al. (2006), ha hecho que se imponga el término **morilla**.

Otro fenómeno que se observa es la aparición de la UTC **olonanacatl**, que permite evidenciar cómo en otras comunidades, para denominar a las especies de *Morchella*, se usan UTC que tienen origen en el recurso metafórico del maíz; provienen del mismo proceso cognitivo y referente cultural, pero el vocablo en sí cambia. En estos casos se encuentra la UTC **mazorca** con su derivativo apreciativo, **mazorquita**; las UTC que provienen del náhuatl, como **olote**, con su sufijo apreciativo **olotitos**, y las UTC que se han tomado como préstamos en el español de México y se han modificado como **elote**, también con su derivativo apreciativo **elotito**.

En el caso del uso de la UTC **mazorca** solo se tiene la variación con su sufijo apreciativo **mazorquita**, y se registra en nueve textos. En el caso de las UTC que vienen de **olotl** o **elotl** hay mayor variación y se registraron cinco, de las cuales tres no se relacionan con derivación apreciativa (**hongo de elote**, **olote** y **elote**). Del uso de estos términos también se debe destacar que no parece haber una selección o límite tan marcado en cuanto al término en las comunidades; de los

tres textos donde aparecen, todos se registran con variación denominativa correspondiente a **mazorca**, **mazorquita** o **morilla**. El fenómeno de variación entre las UTC de pueblos originarios y vocablos del español peninsular en las especies del género *Morchella* es similar a las especies del género *Amanita*, sin embargo, pareciera existir una influencia menor de lenguas originarias en *Morchella*; una de las evidencias puede ser la ausencia de UTC compuestas de palabras con etimología del español peninsular y lenguas originarias como los casos **quismón de huevo** u **hongo de llano**.

Una vez revisados estos datos se puede concluir que la distinción entre las especies de *Morchella* está relacionada con la herencia de la sabiduría cultural de la comunidad, mientras que el término **morilla** se usa para referirse a muchas especies de *Morchella*. Además, aunque *Morchella* sea un género importante para la cultura occidental, y eso aumente su demanda, los pueblos originarios ya lo consideraban un alimento de valor, por lo que sus UTC siguen formando parte de la manera de denominarla.

5 Conclusiones

El resultado de la elaboración del DIGAHONMEX y el análisis de la variación terminológica cultural permitieron un acercamiento a cómo el lenguaje del ámbito gastronómico, específicamente en un país bicultural, no se puede abordar de la misma manera que el lenguaje especializado de las disciplinas u oficios que están solo relacionados con la ciencia. La denominación de los hongos en la gastronomía se comporta de una manera diferen-

te a su denominación científica, pero no por eso deja de ser terminología; solo se trata de una terminología cultural que está íntimamente relacionada con hechos sociales y con dinámicas vivas y comunicativas, por lo que también se reafirma la importancia de la TCT como principio teórico para abordar la terminología cultural.

Así mismo, a partir del planteamiento de la UTC se confirmó que los nombres de los hongos empleados en el área gastronómica sí son términos y que la variación terminológica en el área está causada principalmente por factores funcionales, la necesidad de comunicarse entre recolectores, comerciantes y compradores. Sin embargo, el factor que pareció determinante en el fenómeno de variación es la tensión cultural entre las costumbres occidentales urbanas y la cultura de los pueblos originarios que existe alrededor del uso y la preparación de los hongos en la ciudad. Cada grupo social y cultural intenta conservar las denominaciones propias de sus hongos, aunque, finalmente, quien cede con mayor frecuencia ante el occidentalismo son los pueblos originarios, lo que queda evidenciado en la etimología de las UTC con mayor uso y la alta estandarización en la cocina profesional ante las otras subáreas.

El DIGAHONMEX y su propuesta de macroestructura y microestructura trae al panorama terminológico otras maneras posibles de elaborar diccionarios, atendiendo a los requerimientos de públicos específicos, que muchas veces no han sido considerados como públicos especializados, con lenguajes especializados, al estar relacionados con oficios que no son estrictamente científicos, como es el caso de la gastronomía. ✱

6 Bibliografía

- ALVES DA CONCEIÇÃO COSTA, Lucimara (2015). *Reflexões sobre a variação terminológica na lexicografia corrente no Brasil e a construção das bases teórico-metodológicas para o dicionário de lexicografia brasileira* [en línea]. Tesis doctoral. Universidad Estatal Paulista. Repositorio Institucional UNESP. <<https://repositorio.unesp.br/handle/11449/127683>> [Consulta: 19 junio 2021].
- ARTEAGA MARTÍNEZ, Baldemar; MORENO ZARATE, Clelia (2006). «Los hongos comestibles silvestres de Santa Catarina del Monte, Estado de México». *Revista Chapingo. Serie Ciencias Forestales y del Ambiente*, 12 (2), pp. 125-131.
- AZEVEDO DELVIZIO, Ivanir (2018). «Directrices para la elaboración de definiciones terminológicas: una aplicación a los términos del turismo». *Redillet*, 1.
- BAK-GELLER, Sarah (2013). «Narrativas deleitosas de la nación. Los primeros libros de cocina en México (1830-1890)». *Desacatos*, 43, pp. 31-44.
- BOURET, C. (ed.) (2000). *Nuevo cocinero mexicano en forma de diccionario*. Ciudad de México: Miguel Ángel Porrúa.
- BRODA, Johanna (2019). «La fiesta de Atlcahualo y el paisaje ritual de la cuenca de México». *Trace*, 75, pp. 9-45.
- CABRÉ, M. Teresa (1999). *La terminología: Teoría, metodología, aplicaciones*. Barcelona: Antártida.
- EDO-MARZÁ, Nuria (2012). «Lexicografía especializada y lenguajes de especialidad: fundamentos teóricos y metodológicos para la elaboración de diccionarios especializados». *Lingüística*, 27 (1), pp. 98-114.

- EncicloVida [en línea] (s. f.). <<https://enciclovida.mx/>> [Consulta: 23 junio 2021].
- FAJARDO, José Manuel; VERDE, Alonso; VALDÉS, Arturo; RIVERA, Diego; OBÓN, Concepción (2010). «Etnomicología en Castilla-La Mancha (España)». *Boletín de la Sociedad Micológica de Madrid*, 34, pp. 341-360.
- FAULTISH, Enilde (1995). «Socioterminología: mais que um método de pesquisa, uma Disciplina». *Ciência da Informação*, 24 (3).
- GARIBAY-ORIJEL, Roberto (2009). Los nombres zapotecos de los hongos. *Revista mexicana de micología*, 30, pp.43- 61.
- Guía de Setas y Hongos de Navarra [en línea] <<http://guiahongosnavarrargarciabona.blogspot.com/233>> [Consulta: 28 junio 2021].
- ASOCIACIÓN CULTURAL «BAXAURI» KULTUR ELKARTEA. MIKOLOGIA. BAJAURI. COFRADÍA VASCA DE GASTRONOMÍA (s. f.). Fichas micológicas [en línea]. <<https://www.fichasmicologicas.com/>> [Consulta: 20 junio 2021].
- FREIXA, Judit (2002). *La variació terminològica. Anàlisi de la variació denominativa en textos de diferent grau d'especialització de l'àrea de medi ambient*. Barcelona: Universitat Pompeu Fabra. Institut Universitari de Lingüística Aplicada. (Sèrie Tesis; 3)
- (2013). «Otra vez sobre las causas de la variación denominativa». *Debate Terminológico*, 9, pp. 38-46.
- LARA, Luis Fernando (2016). *Teoría semántica y método lexicográfico*. Ciudad de México: El Colegio de Mexico, AC. (Serie Estudios de Lingüística y Literatura; 66)
- LINARES COLMENARES, Laura (2021). *La variación denominativa de los hongos gastronómicos en México: propuesta de un diccionario gastronómico: Digahonmex*. Tesis de maestría. Universidad Nacional Autónoma de México.
- MORENO FUENTES, Ángel (2013). «Un recurso alimentario de los grupos originarios y mestizos de México: los hongos silvestres». *Anales Antropológicos*, 48 (1), pp. 241-272.
- RUAN-SOTO, Felipe (2007). «50 años de etnomicología en México». *Lacandonia* 1 (1), pp. 97-108.
- UBILLOS, Francisco Javier (s. f.). Fichas micológicas. Asociación Cultural «Baxauri» Kultur Elkartea. Mikologia. Baxauri. Cofradía Vasca de Gastronomía [en línea]. <<https://www.fichasmicologicas.com/>> [Consulta: 20 junio 2021].

Nota:

1. Cocina rural: área en la que un cocinero o grupo de cocineros ejecuta preparaciones para comensales íntimos, en un contexto semirrural, rural o en comunidades originarias. Los ingredientes empleados se obtienen de sus propias cosechas, de las cosechas de los vecinos y centros de compras medianos (abarrotes o mercados itinerantes). Los cocineros no suelen tener escolaridad formal, sin embargo, usan la sabiduría local y familiar para desarrollar sus preparaciones. En la mayoría de los casos estas preparaciones cumplen el objetivo de cubrir las necesidades diarias de la alimentación de su familia, pero también pueden tener fines recreativos y sociales. En algunas ocasiones pueden vender sus preparaciones para mejorar sus ingresos.

Cocina doméstica: área en la que un cocinero o grupo de cocineros ejecuta preparaciones para comensales íntimos, en un contexto urbano. Los ingredientes empleados se obtienen de centros de compras industrializados o centros de compras medianos y se tiene acceso a ingredientes importados, locales y orgánicos. Los cocineros pueden tener formación profesional o no; en algunos casos puede ser la fuente de ingresos si el cocinero trabaja para una familia que no es la suya. En la mayoría de los casos las preparaciones tienen el objetivo de cubrir las necesidades diarias de alimentación, pero también pueden emplearse con fines recreativos y sociales.

Cocina profesional: área en la que un cocinero o grupo de cocineros ejecutan preparaciones para comensales que pagan por este servicio en un centro de consumo rural o urbano. Los ingredientes se obtienen de proveedores, centros de compras industrializados y se tiene acceso a ingredientes importados, locales y orgánicos. En algunos casos usan productos de sus propias cosechas. Los cocineros pueden tener formación académica o haber adquirido los conocimientos del oficio a través de experiencia o algún vínculo familiar. El trabajo es remunerado y generalmente es el ingreso principal del cocinero.

Una mirada dialectològica al Observatorio de Palabras de la RAE vía Twitter

JULIA ELENA PARDO

Universitat Pompeu Fabra

jpardo.al@gmail.com

Julia Elena Pardo és graduada en

Enginyeria per la Universitat de Buenos Aires i màster en Lingüística Teòrica i Aplicada per la Universitat Pompeu Fabra. El seu treball de fi de màster sobre dialectologia computacional i prescriptivisme popular va brindar les bases metodològiques per a aquest treball de terminologia. Els seus interessos se centren en la variació lingüística i la lingüística popular, usant la ciència de dades com a eina principal.



Resumen

El Observatorio de Palabras recoge términos —fundamentalmente, neológicos— que no aparecen en el Diccionario de la lengua española y que han generado dudas de los hispanohablantes a la Real Academia Española (RAE). Su función normativa, sin embargo, se presta a confusión. Con el objeto de esclarecerla, este artículo toma algunos de estos términos y analiza la evolución de su uso y su variación diatópica en el ámbito panhispánico, utilizando Twitter como corpus dialectológico. Se satisfacen los criterios de neologización y se sugiere el acercamiento de la Academia al hablante. PALABRAS CLAVE: RAE; Twitter; neología; dialectología panhispánica

Resum

Una mirada dialectològica a l'Observatorio de Palabras de la RAE via Twitter

L'Observatorio de Palabras recull termes —fonamentalment, neològics— que no apareixen en el Diccionario de la lengua española i que han generat dubtes dels hispanoparlants a la Real Academia Española (RAE). Tanmateix, la seva funció normativa porta a confusió. A fi d'esclarir-la, aquest article pren alguns d'aquests termes i n'analitza l'evolució pel que fa a l'ús i la variació diatòpica en l'àmbit panhispànic, utilitzant Twitter com a corpus dialectològic. Se satisfan els criteris de neologicitat i se suggereix l'acostament de l'Acadèmia al parlant.

PARAULES CLAU: RAE; Twitter; neologia; dialectologia panhispànica

Abstract

A dialectological look at RAE's Observatorio de Palabras on the basis of Twitter

The Observatorio de Palabras («Word Observatory») collects terms — chiefly neologisms — that are not in the Diccionario de la lengua española («Dictionary of the Spanish language») and that have led Spanish speakers to convey their respective doubts to the Real Academia Española (Royal Spanish Academy, RAE). The Word Observatory's normative role, however, gives rise to confusion. Aiming to clarify it, this article takes some of these terms and analyses the evolution of their use and their diatopic variation at a pan-Hispanic level, using Twitter as a dialectological corpus. Neologicity criteria are met and a closer approach of the Academy to the speakers of the language is suggested.

KEYWORDS: Twitter; neology; pan-Hispanic dialectology

TERMINÀLIA 27 (2023): 32-50 · DOI: 10.2436/20.2503.01.187

Data de recepció: 19/10/2022. Data d'acceptació: 05/04/2023

ISSN: 2013-6692 (impresa); 2013-6706 (electrònica) · <http://terminalia.iec.cat>

1 Introducción

El Observatorio de Palabras (OdP) fue presentado por la Real Academia Española (RAE) en su página web el 23 de octubre de 2020. Allí se describía de la siguiente manera (RAE, 2020a):

El «Observatorio de palabras» ofrece información sobre palabras (o acepciones de palabras) y expresiones que no aparecen en el diccionario, pero que han generado dudas: neologismos recientes, extranjerismos, tecnicismos, regionalismos, etc. Esta información es provisional, pues no está contemplada en las obras académicas, por lo que puede verse modificada en el futuro. La presencia de un término en este observatorio no implica que la RAE acepte su uso.

En tan solo unos pocos días, era noticia en todos los países de habla hispana: el pronombre *elle* se encontraba entre el conjunto de palabras descritas, y la confusión generalizada sobre su grado de aceptación y su estado de evaluación por parte de la RAE inundó la prensa de todas las ideologías. Tal fue la repercusión que la RAE pasó —de inmediato— a eliminar la entrada del controvertido pronombre. Alegó que se volvería a valorar su inclusión «cuando se difundiera ampliamente el funcionamiento y cometido de esta sección» (RAE, 2020b). Hoy día, casi dos años más tarde, la descripción permanece intacta (RAE, 2022). Sabemos, por respuestas que ha dado la RAE en Twitter a usuarios que consultaban, que las setenta y tres palabras que allí aparecen fueron incluidas en el 2020 (RAE, 2021a), y que su objetivo «no es recoger las palabras que están en fase de incorporación al diccionario, sino resolver dudas relacionadas con algunas de las palabras ausentes del DLE» (RAE, 2021b).

La introducción del OdP representa una innovación dentro de los procedimientos habituales de la RAE; sin embargo, el reclamo de una mejor especificación de su cometido y una adecuada distinción académica entre neologismos y extranjerismos, tecnicismos y localismos permanece vigente (García Platero, 2022, p. 29). En efecto, las entradas del OdP no poseen clasificación alguna. Más aún —favoreciendo la confusión general—, algunas de ellas pueden encontrarse en el *Diccionario de la lengua española* (DLE) (RAE y ASALE, 2022), en el *Diccionario panhispánico de dudas* (DPD) (RAE y ASALE, 2005) y en el *Diccionario de americanismos* (DA) (RAE y ASALE, 2010). En el OdP, la descripción de cada entrada detalla si son o no voces válidas, qué uso o usos son los sugeridos o recomendados, y cuáles se desaconsejan, sea por tratarse de extranjerismos, sea por ser considerados innecesarios, al existir otra variante en español. Esto último acarrea la inquietud respecto al prestigio de la variante española preferida: ¿se trata de la variante globalmente preponderante o es propia de ciertas regiones? ¿Cómo se condicen las preferencias y sugerencias que da la RAE con los usos y tendencias populares?

Lo cierto es que las palabras que conforman el OdP han generado dudas entre los hablantes del idioma y, por ello, se encuentran compiladas en este portal lingüístico. Estos hablantes pertenecen a una comunidad de casi seiscientos millones de personas (Instituto Cervantes, 2022, p. 23), geográficamente extensa y rica en variación dialectal. El español es concebido actualmente como una lengua pluricéntrica: existen muchos estándares de la lengua, con diferente alcance regional, pero en su mayoría carecen de apoyo institucional, de una codificación formal (Amorós-Negre, 2018, pp. 57-58). Estas variedades ejemplares constituyen gradaciones no solo desde el punto de vista diatópico, sino también diastrático y diafásico en el plano comunicativo. La norma española vigente, en tanto, describe y prescribe un estándar composicional (Senz et al., 2011, § 3.5.4), una única variedad formalmente codificada, entendida como una «lengua culta general panhispánica» (Amorós-Negre, 2018, p. 63). Como tal, tiene un carácter supradialectal que garantiza la comunicación entre la extensa comunidad de hispanohablantes (Amorós-Negre, 2018, pp. 48-49).

La actual norma española es el reflejo de una política lingüística reciente. Si bien la Asociación de Academias de la Lengua Española (ASALE) fue fundada en 1951 con el objeto de iniciar una política colaborativa entre todas las academias de la lengua (Paffey, 2021, p. 244), no fue hasta finales de siglo que se produjo una colaboración genuina (Amorós-Negre, 2018, p. 60) y hasta 2004 que se materializó en la «nueva política lingüística panhispánica». A la fecha, la ASALE está integrada por veintitrés academias de América, España, Filipinas y Guinea Ecuatorial (ASALE, 2022), pero es la RAE quien ha logrado mantener y naturalizar su papel público como autoridad de referencia de la lengua, representativa de todo el mundo hispanohablante (Paffey, 2021, p. 250; Valle, 2007, p. 88). Es a ella a quien el hablante común acerca sus dudas, y Twitter es uno de los medios de interacción, donde —desde 2012, como @RAEinforma— la RAE responde consultas y recibe todo tipo de opiniones y sugerencias.

En efecto, para el hablante común el buen lenguaje no es el reflejo de una variante prestigiosa, elegida y expresada para un estándar normativo, se trata de un ente más abstracto, que es bueno porque se presupone más lógico, claro y etimológicamente continuo (Niedzielski y Preston, 2000, p. 18). Así, acude a diccionarios y gramáticas como una incursión a tal abstracción del lenguaje y no como a meras autoridades sobre su uso. Las obras normativas trascienden su carácter descriptivo y se establecen como autoridad prescriptiva; por ello, la entrada de una palabra al diccionario le confiere un aval de garantía y de prestigio (Parakrama, 1995, pp. 10-12 y 28-29; Prieto de los Mozos, 1999, p. 249; Guerrero Ramos, 2022, pp. 191-193). Sin embargo, el OdP no es un diccionario, ni es parte de la norma; tampoco es necesariamente un estadio inter-

medio (aunque lo haya sido en algunos casos). Pero es natural que, siendo este portal lingüístico presentado por la autoridad prescriptiva, haya generado confusión en cuanto a la validación que concede a sus entradas.

Pero ¿por qué estos términos agrupados en el ODP han sido cuestionados? ¿Cómo ha sido su uso en los últimos años? ¿Se observan diferencias entre los distintos países de habla hispana? ¿Podrían algunos de estos términos ser incorporados al diccionario? ¿Qué características o comportamientos encontramos para los ya aceptados?

El objetivo de este trabajo es, tomando como base una selección de las entradas del ODP y utilizando Twitter como corpus dialectológico (véase § 3), establecer y analizar la evolución de sus frecuencias de uso y su distribución diatópica en los distintos países hispanohablantes; sean o no estos términos eventualmente aceptados, se busca compararlos con los que ya lo han sido (en la última edición electrónica del DLE) y observar cómo se ajustan las recomendaciones de la RAE a los usos de la gente corriente en su hablar cotidiano. Se pretende, en definitiva, esclarecer la función que cumple este observatorio.

2 Las nuevas voces del español

La variación y los cambios lingüísticos son inherentes a cualquier lengua viva; la reacción a estos cambios por parte de los hablantes también es natural y se encuentran en tensión permanente la conservación y la innovación del idioma (Jiménez Ríos, 2015, p. 46). Para que un cambio dado logre estabilidad y sea, posteriormente, incorporado a gramáticas y diccionarios, requiere que se alcance sobre él un consenso social, que se manifiesta tanto en la frecuencia de uso como en su perdurabilidad (Lorente Casafont, 2022, p. 35). Pero, tal como lo indica Guerrero Ramos (2022), aun si la neologidad de una voz puede medirse con cierta objetividad según el criterio lexicográfico, su incorporación al diccionario depende tanto de la aptitud como de la ideología del equipo de redacción involucrado (p. 191). En efecto, las opiniones de los lexicógrafos respecto a la idoneidad o conveniencia de utilizar ciertos neologismos —en particular, la discriminación entre lo necesario y lo superfluo— difícilmente están alineadas (p. 192). Y el hablante —incluso a pesar de su naturaleza creativa— recurre al diccionario (o a la autoridad) en busca de la validación del uso de estas nuevas voces (p. 209). Al igual que los expertos, los hablantes difieren en sus opiniones, y tales discrepancias, sus expresiones de acuerdo y desacuerdo para con la RAE y entre ellos mismos, quedan manifestadas en los hilos públicos de Twitter de @RAEinforma, por lo que intervienen en la regulación de los usos, con prácticas prescriptivas populares o de base (Lukač, 2018, p. 5).

La variación del español —diatópica, diastrática y diafásica— se refleja, naturalmente, también en la

neología. Así, se encuentran voces nuevas comunes a las distintas variedades, otras propias a ciertas regiones y también voces nuevas distintas que responden a un mismo significado (Fernández-Silva, 2022, p. 259). Mientras la distancia geográfica y las diferencias culturales, sociopolíticas y económicas entre las comunidades influyen en la diferenciación entre las lenguas locales, factores externos como los medios masivos, la urbanización, la movilidad global y las redes sociales favorecen la interacción lingüística de los hablantes (Amorós-Negre, 2018, p. 65; Sánchez Lobato, 2019, pp. 68-69). Son ellos, en última instancia, los responsables de facilitar y preservar su comprensión mutua (Lara, 2015, pp. 27-28, citado en Amorós-Negre, 2018, p. 65), valorando sus similitudes y peculiaridades lingüísticas (Sánchez Lobato, 2019, p. 68). La innovación lingüística del español se encuentra, pues, en una tensión entre los procesos de variación, propiciados por los fenómenos locales, y los de homogeneización, fomentados por las dinámicas globales (Fernández-Silva, 2022, p. 262). Las interacciones en Twitter entre los hispanohablantes y la RAE reflejan estas tensiones, así como las variadas actitudes y opiniones respecto a las nuevas voces: «consulta» (Carla P., 2014), «sugerencia» (Arnandis, 2017), «aprobación» (Negrete-Portillo, 2018), «crítica» (exYge, 2020), «rechazo» (Daluna MX, 2018) y «sorpresa ante usos regionales sugeridos» (Felipe el Iguano, 2019).

Las obras normativas, por un lado, favorecen la homogeneización en la presentación de una norma común panhispánica que, simultáneamente, procura reconocer el pluricentrismo empírico de sus variedades. A pesar de las interpretaciones divergentes que genera el trasfondo ideológico de la política lingüística, el carácter compositivo del estándar actual se ha apartado de sus antiguas formas eurocéntricas para acercarse a las de sus hablantes (Greußlich, 2015, pp. 59-60; Amorós-Negre, 2018, p. 308). Sin embargo, aún es notorio el tratamiento diferenciado de las variantes no peninsulares como subsidiarias o marcadas, pues sigue vigente la distinción centralista entre el «diccionario general» (el DLE) y los «diccionarios de regionalismos», en particular, el DA (Lara, 2005, p. 182). Por un lado, estos diccionarios diferenciales o de -ismos solo incluyen lo disímil y periférico, no son autosuficientes y apelan a un público especializado y no al general; por el otro, hay aún una falta de representación de estos usos «marcados» en los diccionarios generales, que pretenden abarcar la lengua en su totalidad (Ávila, 2004, pp. 13-14; Fernández-Silva, 2022, pp. 259-260). Similarmente, los diccionarios de neologismos tampoco han mostrado un tratamiento prioritario hacia el eje diatópico. No obstante, tratándose el español de una lengua de notable expansión geográfica e histórica, con gran vitalidad léxica, la variación diatópica no puede ser desatendida, pues tiene inevitable influencia en las formas y las motivaciones de la producción neológica (Fernández-Silva, 2022, pp. 260-261).

Como observa Freixa (2022) en su estudio del proceso neológico, la institucionalización de una voz —entendida como la difusión y establecimiento social de su uso— no necesariamente culmina con su incorporación al diccionario (p. 57), ya que los criterios lexicográficos pueden primar sobre la evolución en su uso (p. 71). Por ello, Freixa (2022) sugiere que la solución más apropiada es analizar frecuencias de uso en grandes corpus de textos de la lengua general, por ejemplo, mediante histogramas (p. 58). Estos revelan distintos patrones en la evolución de neologismos; así por ejemplo, los candidatos a incorporarse al diccionario suelen compartir cinco características: no tender al desuso, estar en uso desde hace al menos cuatro años, tener más de diez ocurrencias, aparecer en variados tipos de textos y hacerlo en textos provenientes de variedades geolectales diferentes (pp. 64-65).

Cabe notar también la influencia de la variación diafásica en los neologismos. Como indica Sanmartín (2022), estos se originan y utilizan en distintos ámbitos comunicativos y discursivos, entre los que destacan, sobre todo, el léxico de especialidad y el léxico coloquial (pp. 234-236). La prensa, que es una fuente habitual para el estudio de neologismos, suele tener un registro más formal, aunque en ocasiones recurre a usos informales como recurso estilístico (pp. 241 y 257). Si bien, en el plano léxico, es más sencillo dar a una nueva voz una marca coloquial o neutra que una formal, debe verificarse como tal en un corpus coloquial o en redes sociales (p. 257).

3 Twitter como corpus dialectológico

El estudio computacional sobre cambio y variación lingüística mediante el uso de las redes sociales ha tenido un gran crecimiento en los últimos años (véase la revisión realizada por Nguyen *et al.* [2016] sobre sociolingüística computacional). Dado que el medio de comunicación no necesariamente define la concepción oral o escrita (Koch y Oesterreicher, 2007), el lenguaje de las redes sociales —a pesar de tratarse de un medio gráfico— puede presentar numerosas características de la oralidad, con lo que consigue gran heterogeneidad y es sumamente útil para el estudio variacional (Estrada Arráez y Benito Moreno, 2016, § 2). Twitter, en particular, por su extensión limitada, su naturaleza pública (aunque informal) no necesariamente conversacional y su exigua regulación —que concede absoluta libertad temática—, constituye su propia variedad, un híbrido entre las diferentes variedades de géneros en línea (Jehl, 2010, § 3.3). Adicionalmente, Twitter proporciona datos geolocalizados precisos, lo que convierte los tuits geoetiquetados en la fuente de numerosos estudios que han favorecido el mapeo de la variación lingüística (Grieve *et al.*, 2019, p. 1). Entre aquellos sobre el español global, Gonçalves y Sánchez (2014 y 2016) analizan la variación léxica y establecen

grupos dialectales, Ruiz Tinoco (2016) compara distintos usos léxicos y gramaticales en España y Latinoamérica, y Pardo (2022) analiza usos frecuentemente cuestionados por los hispanohablantes.

A pesar del alcance ilimitado en temática, proveniencia geográfica y número de informantes (Ruiz Tinoco, 2016, p. 599; Grieve *et al.*, 2019, p. 15), predominan las cohortes jóvenes, masculinas y urbanas (Longley *et al.*, 2015, § 3.1; Eisenstein, 2017, p. 370), por lo que la representatividad demográfica no es completamente equilibrada. Incluso así, varios estudios han ratificado a Twitter como una fuente válida para componer mapas dialectales (Cook *et al.*, 2014; Doyle, 2014; Huang *et al.*, 2016) y estudiar patrones generales de variación léxica regional (Grieve *et al.*, 2019). Lamentablemente, las restricciones legales de intercambio de datos dificultan la creación de corpus públicos, lo que afecta la reproducibilidad de los resultados, pues están sujetos a ser borrados (Ruiz Tinoco, 2016, § 2.1; Estrada Arráez y Benito Moreno, 2016, pp. 96 y 106). A pesar de la facilidad de recolectar muestras de gran tamaño (Huang *et al.*, 2016, p. 254; Grieve *et al.*, 2019, p. 15), la creación de corpus representativos requiere mucho tiempo de relevo, en especial cuando involucra a países de distinto producto interno bruto (PIB), pues varía su producción de tuits (Dunn y Adams, 2020, p. 4). En consecuencia, resulta imperante utilizar directamente el motor de búsqueda de la interfaz de programación de Twitter, incluso con la limitación de que —además de no reconocer diacríticos— no permite el uso de expresiones regulares (secuencias de símbolos y caracteres que expresan una cadena o patrón que pueda ser buscado dentro de un texto más extenso), lo que hace del proceso un trabajo más artesanal.

Si bien los tuits geoetiquetados representan solo el 1-2 % del total (Huang *et al.*, 2016, p. 253) y responden a un subgrupo sociodemográfico (Malik *et al.*, 2015), también es posible utilizar la ubicación en texto libre que provee el usuario en su perfil. A pesar de que depende de la voluntad del usuario de informar su ubicación real, Pavalanathan y Eisenstein (2015) encontraron que en la mayoría de los casos son precisas (§ 3.1) y que las ubicaciones geoetiquetadas y autoinformadas producen corpus diferentes (§ 4.2). Considerar ambas crea un conjunto de datos mayor que puede generar resultados más confiables (Huang *et al.*, 2016, p. 253).

4 Metodología

Con el objetivo de determinar las frecuencias de uso y la variación diatópica de las entradas del OdP, estas fueron seleccionadas con el criterio de poder establecer pares de variantes de alternancia (palabras alternativas para un mismo significado). El estudio de la frecuencia relativa de términos en el ámbito del país resulta fundamental para obtener resultados fiables en el caso de términos muy poco frecuentes, evalua-

dos en países cuya producción de tuits es muy dispar. En algunos casos, las entradas del Observatorio tienen un equivalente preexistente en el DLE, que se usó para constituir el par, y, en otros, son las propias entradas las que conforman el par de alternancia (con o sin equivalentes aceptados). Se desestimaron entradas polisémicas, a excepción de aquellas con un uso secundario muy infrecuente, en cuyo caso ese uso fue desestimado para el análisis (p. ej.: *dilogía*).

Así, de las 73 palabras del ODP, se seleccionaron 23 y se definieron un total de 13 pares de variantes de alternancia (algunas colectivas), presentados en la tabla 1. Allí se incluye una breve reseña de lo que indica la RAE en el ODP para cada palabra (RAE, 2022) y, en caso de ya encontrarse en alguno de los diccionarios DLE, DA y DPD, se recogen estos datos.

El proceso de extracción y análisis de datos se llevó a cabo usando Python 3.10.1 (Rossum y Drake, 2009) y sus bibliotecas de análisis de datos, así como la interfaz de programación de Twitter (<https://developer.twitter.com/>).

La frecuencia absoluta de uso de cada variante —siguiendo a Gonçalves y Sánchez (2014) y Ruiz Tinocho (2016)— se determinó por la cantidad de tuits que incluían dicha variante. Se incluyeron filtros para recolectar tuits en español y para desestimar retuits, con lo que se evitaron textos replicados. Según fuera posible, se contemplaron variación en género y número, conjugación en presente de indicativo, sustantivos, adjetivos, verbos y participios derivados. En algunos casos, se analizaron diferencias de uso entre verbos y sustantivos derivados de una misma entrada, bajo la hipótesis de que la conjugación verbal podría ser favorable a una versión ortográficamente más castellanizada de los extranjerismos.

La población de estudio son los usuarios hispanohablantes de Twitter, pero —siguiendo a Pardo (2022, § 8)— solo fueron considerados aquellos países cuyas academias pertenecen a la ASALE. Para determinar la ubicación de cada tuit, se tomó en cuenta su geolocalización y el campo de ubicación informado por el usuario en su perfil (rango de error de 0,4 a 1,1 % con un 95 % de confianza), en ese orden de prioridad y cuando estuvieran disponibles (Pardo, 2022, § 8.4). Esto permite extraer más tuits válidos (con una ubicación identificable), conveniente especialmente para palabras poco frecuentes y para lograr una mejor representatividad demográfica.

El análisis diatópico se realizó empleando frecuencias de uso relativas, utilizando cada país como unidad regional, esto es, la frecuencia de uso (cantidad de tuits) de cada variante con respecto a la de ambas variantes del par, por país (Pardo, 2022, § 8.6):

$$f. \text{ rel. var1} | \text{país}_i = \frac{\text{tuits. var1}}{\text{tuits. var1} + \text{tuits. var2}} \Bigg|_{\text{país}_i}$$

Por ello, los tuits fueron recolectados por par de alternancia: los tuits que contenían cualquiera de las variantes se buscaron y recopilamos simultáneamente durante el mismo período de tiempo, con lo que se aseguraba la proporción de uso apropiada de cada variante para el número total de tuits descargados. El periodo de recolección fue de enero de 2010 a julio de 2022, a excepción de usos muy frecuentes en los que el intervalo fue menor (siempre desde 2022 hacia atrás). Estos datos fueron luego procesados para determinar su proveniencia y utilizados para el mapeo dialectológico. Este método permite independizarse de la diferencia de producción de tuits entre países y ver claramente cómo se distribuyen los usos regionalmente. En pocos casos, en los que hay más de dos variantes alternantes, se creó una variante colectiva para poder establecer un mapeo adecuado, se agruparon las variantes poco frecuentes en un mismo conjunto y se contrastó su uso relativo respecto a la preponderante: se agrupó las variantes poco frecuentes en un mismo conjunto y se las contrastó en uso relativo respecto a la preponderante (p. ej.: *ranking* vs. *ranquing*, *ranquín*, *rankin*).

Los pares de alternancia se organizaron en dos grupos ordenados consecutivamente (en la tabla 1: 1 a 6 y 7 a 13), según las posibilidades de su análisis temporal, esto es, según el volumen de datos disponible para la recolección. En los primeros seis pares, ambas variantes del par (o una variante colectiva) permitieron la recolección de tuits desde 2010 hasta julio de 2022, lo que posibilitó un análisis temporal comparativo; en los siguientes siete pares, solo la variante minoritaria pudo recolectarse para ese mismo período, por lo que su análisis temporal fue individual (tras la recolección simultánea para el análisis diatópico, se prosiguió por separado en la recolección hasta 2010). En todos los casos, el número de tuits de cada año, para cada variante, fue ponderado por la cantidad de usuarios de Twitter del año respectivo a nivel global, según datos de la web Statista (Statista, 2022a y 2022b), de modo que se reflejasen las variaciones en el uso y no las variaciones en cantidad de usuarios (número significativo en los inicios de la red social). Si bien los factores de ponderación responden a números de usuarios globales y no hispanohablantes en particular, fue considerado un compromiso necesario.

Para ambos análisis se descartaron los tuits que incluyeran ambas variantes o que mencionaran a la RAE, ya que esto podría implicar una reflexión de los usuarios sobre el idioma, en lugar de una producción espontánea (la alternancia inconsciente se descartó en favor de lo anterior). Se exigió un mínimo de veinte tuits por país para incluir el par en el análisis; de esta forma, la probabilidad de una asignación incorrecta en la ubicación representaría únicamente un 5 % del total.

Tanto en el caso de *dilogía* como *destripe* (uso peninsular para *spoiler*), su polisemia fue desestimada para el análisis.

N.º par	Entrada del OdP	Descripción del OdP + Detalle en obras normativas	Par de alternancia
1	eficientizar	Verbo válido, de empleo habitual en varios países de América. + En el DA: Nicaragua y República Dominicana.	eficientizar vs. eficientar
	eficientar	Verbo válido, de empleo habitual en varios países de América. + En el DA: Nicaragua y Honduras.	
2	gamificación	Adaptación híbrida desaconsejada de <i>gamification</i> . Se recomienda «ludificación».	gamificar vs. ludificar
	ludificación	Equivalente recomendado a <i>gamification</i> . Propuesta para el DLE. + En el DLE desde diciembre del 2021.	
3	turistólogo	Voz reciente, en alternancia con «turismólogo».	turistólogo vs. turismólogo
	turismólogo	Voz reciente, en alternancia con «turistólogo».	
4	dilogía	Voz correcta, en alternancia con «bilogía». Se recomienda «obra de dos partes». + En el DLE, con otro significado.	dilogía vs. bilogía
	bilogía	Voz correcta, en alternancia con «dilogía». Se recomienda «obra de dos partes».	
5	ranquin	Adaptación recomendada para <i>ranking</i> . Propuesta para el DLE. + En el DPD.	ranking vs. ranquin, ranquing, rankin
6	webinar	Anglicismo. Se recomiendan: «webinario», «seminario web» o «ciberseminario». + «webinario» y «seminario web» en el DLE y «ciberseminario» se considera un compuesto a partir de «ciber-» (también en el DLE).	webinar vs. webinar, seminario web, ciberseminario
7	mutear	Derivado español, a partir de <i>mute</i> . Sustituir por «silenciar».	mutear vs. silenciar
8	legitimizar	Voz innecesaria y desaconsejada. Sustituir por «legitimar».	legitimizar vs. legitimar
9	spoiler	Anglicismo. Sustituir por «destripe» o «espóiler».	spoiler vs. espóiler
	espóiler	Adaptación documentada y válida.	
	destripe	Derivado de «destrigar». Alternativa a <i>spoiler</i> .	
10	guasap	Adaptación válida.	guasap(ear) vs. wasap(ear)
	wasap	Adaptación válida.	
	guasapear	Grafía adecuada y documentada.	
	wasapear	Grafía adecuada y documentada.	
11	guglear	Grafía adaptada y documentada.	guglear vs. googlear
	googlear	Grafía híbrida y desaconsejable.	
12	emoyi	Posible adaptación de «emoji», grafía preferida. + «emoji» en el DLE desde diciembre del 2021.	emoyi vs. emoji
13	paranoya	Grafía aceptable de «paranoia», grafía tradicional.	paranoya vs. paranoia

TABLA I. Entradas seleccionadas del OdP y par de alternancia asociado
Todas las tablas son de elaboración propia. En este caso, las fuentes de información son:
RAE, 2022; RAE y ASALE 2022, 2010, 2005 (DLE, DA y DPD, respectivamente)

5 Resultados

La cantidad de tuits recolectados fue muy variable en función del par de alternancia: se obtuvieron una media y una desviación estándar globales de (184 ± 120) mil tuits. El par más escaso fue *turistólogo/turismólogo*, con solo 8.361 tuits desde 2010. El requerimiento de un mínimo de veinte tuits por país limitó la inclusión de Guinea Ecuatorial en solo dos pares de alternancia y a la mitad en el caso de Filipinas. Sin embargo, en todos los análisis de alternancia se contó con la mayoría de los países hispanohablantes.

La tabla 2 resume los resultados de los análisis temporal y diatópico de cada par alternante. Para el análisis temporal, se clasifica el estado de evolución actual siguiendo a Freixa (2022, § 5); para el análisis diatópico, se presentan las distribuciones de uso encontradas para la variante minoritaria globalmente (no necesariamente localmente); se incluyen la media y la desviación estándar (SD) de su frecuencia relativa de uso a nivel global, así como el detalle de los tres países que mostraron un mayor uso (estos valores son complementarios a los de la variante mayoritaria; la SD expresa la dispersión de los datos respecto a la media). Tanto las gráficas temporales como los mapas dialectológicos se encuentran disponibles en el anexo.

El par *eficientar/eficientizar* presentó, para ambos casos, un uso estabilizado en el mundo hispanohablante, a pesar de diferencias en amplitud debido a la disparidad de uso de ambas variantes, con clara preponderancia a favor de *eficientar*; el análisis diatópico confirmó su preeminencia a nivel global (5,9 %) y local, aunque en República Dominicana alterna casi en igual medida con *eficientizar*. A pesar de que ambos términos figuran en el DA, sendos usos aparecen con la marca diatópica de Nicaragua; *eficientizar* con la de República Dominicana, y *eficientar* con la de Honduras. Sin embargo, en Nicaragua no se alcanzó el mínimo requerido de veinte tuits entre ambas variantes y, aunque se trata de un par poco frecuente (33.876 tuits), hubo otros pares alternantes de frecuencia similar para los que Nicaragua sí satisfizo el mínimo. Argentina es el otro país donde *eficientizar* tiene un uso relativamente significativo, un 35 %, seguido de Colombia, con tan solo el 4 %. España no quedó excluida del análisis (puesto diez de dieciséis, con un 0,2 % de uso), por lo que puede afirmarse que este uso no se encuentra restringido a América. Los resultados, para este par alternante, no se corresponden con lo que sugiere la norma académica vigente. Comparando este ejemplo con el par *legitimar/legitimizar*, donde podría esperarse algún paralelismo en el uso, se observa cierta estabilidad general en la variante minoritaria, que tiene una media de uso similar (6,3 %), pero una desviación estándar menor y una distribución distinta entre países, aunque los de mayor uso también son americanos. La RAE, en este caso, no recomienda su uso, lo que está en consonancia con la preferencia popular.

El par *gamificar/ludificar* muestra una clara preponderancia del término desaconsejado por la RAE. Mientras que *ludificar* alcanza, globalmente, apenas un 2,8 % de uso (SD 1,5 %), este fue incluido en el DLE en 2021 (seguramente, dada su etimología). Entre las críticas de los usuarios se encuentran el retraso de la RAE en su aporte y la falta de asociación del término propuesto con su uso (Lara, 2019; exYge, 2020).

El caso del par *turistólogo/turismólogo* evidencia un patrón neológico realista (ascenso, descenso y estabilización), aunque la preferencia a nivel global parece haberse invertido en favor de *turismólogo*. Sin embargo, el término minoritario a nivel global no lo es a nivel local, con una media de 37,8 % y una SD de 23,7 %, siendo el par que exhibió más variabilidad regional.

Bilología ha primado sobre *dilogía*, mostrando la preferencia por el prefijo latino «bi-» sobre el griego «di-». Incluso habiendo desestimado el significado presente en el DLE, el uso de *dilogía* es notablemente minoritario (3,5 %); destaca Cuba con el 20,8 %.

Los casos de los anglicismos *ranking*, *webinar* y *spoiler* muestran que estos son ampliamente preponderantes sobre sus respectivas variantes en español. Dentro de las minoritarias, se observa el uso mayoritario de *rankin* sobre las variantes con «q» (*ranquin* está incluido en el DPD), y la preferencia de *webinario* y *seminario web* sobre un casi inexistente *ciberseminario*, todos ellos con un uso acentuado en la pandemia de 2020, pero respetando las tendencias previas. Notoriamente, estas versiones minoritarias en español (a excepción de *espóiler*, que ronda tan solo el 0,5 % de uso) tienen un uso más elevado en Estados Unidos, quizás motivado por un deseo de mayor distinción respecto al inglés. Es similar lo que ocurre con los pares *wasap/guasap* y *googlear/guglear*, donde, si bien se trata de versiones en español, la primera —más cercana a la inglesa— es la más usada; el pico temporal de *guasap* coincide con la popularización masiva de la aplicación Whatsapp, mientras que *guglear* ya se muestra estabilizada.

En todos los casos en los que se estudió la diferenciación entre sustantivos y verbos, se observó, para los últimos, un incremento notable en el uso de la variante de ortografía castellanizada (en especial, *ranquear* y *espoilear*, menor en *guasapear*).

En los pares *emoji/emoj* y *paranoya/paranoia* el uso con «y» es insignificante, prácticamente inexistente en conjunto en el caso de *emoji* y, para *paranoya*, se destaca únicamente España con un 6,4 % de uso. Mientras *paranoya* parece haber estabilizado su uso, restaría ver la evolución temporal de *emoji* en el futuro. Es llamativo que ambos usos, tan infrecuentes, sean sugeridos como válidos por la RAE.

Finalmente, cabe aludir al caso de *mutear*, del que, en competencia con *silenciar*, cabría esperar un uso muy limitado, pero que alcanzó un 8,1 % global y superó el 20 % de uso en Puerto Rico y República Dominicana. El pico de frecuencia se vio favorecido por la pandemia. De hecho, los usuarios se dirigieron a

N.º par	Var. minoritaria (var. mayoritaria)	media ±SD	Análisis diatópico		Mapa	Análisis temporal
			Países principales			
1	eficientizar (eficientar)	5,9 % ±13,9 %	R. Dominicana Argentina Colombia	47,5 % 34,3 % 4,1 %		 estabilización
2	ludificar (gamificar)	2,8 % ±1,5 %	Bolivia Cuba Uruguay	7,4 % 5,5 % 3,8 %		 ascenso y descenso
3	turismólogo (turistólogo)	37,8 % ±23,7 %	Perú España Ecuador	79,7 % 76,0 % 61,5 %		 ascenso, descenso y estabilización
4	dilogía (bilogía)	3,5 % ±4,7 %	Cuba Bolivia Perú	20,8 % 8,3 % 8,1 %		 ascenso, descenso y estabilización
5	ranquin, rankin, ranquing, (ranking)	1,5 % ±1,4 %	Méjico EE. UU. Puerto Rico	5,6 % 3,5 % 3,2 %		 descenso y estabilización
	ranquear (rankear)	15,7 % ±8,5 %	R. Dominicana Puerto Rico Colombia	36,7 % 31,2 % 23,4 %		 descenso y estabilización
6	webinario, seminario web, ciberseminario (webinar)	9,5 % ±4,7 %	EE. UU. El Salvador Uruguay	19,7 % 19,3 % 13,6 %		 ascenso y descenso
7	mutear (silenciar)	8,1 % ±5,5 %	Puerto Rico R. Dominicana Costa Rica	22,9 % 20,7 % 17,8 %		 ascenso
8	legitimizar (legitimar)	6,3 % ±4,3 %	Puerto Rico Panamá Bolivia	20,2 % 13,0 % 12,7 %		 estabilización

Análisis diatópico						
N.º par	Var. minoritaria (var. mayoritaria)	media ±SD	Países principales		Mapa	Análisis temporal
9	espóiler (spoiler)	0,5 % ±0,2 %	Cuba Panamá España	1,0 % 0,8 % 0,7 %		 descenso y estabilización
	espoilear (spoilear)	3,4 % ±1,7 %	Puerto Rico Chile R. Dominicana	9,4 % 5,1 % 4,9 %		
10	guasap (wasap)	13,2 % ±9,1 %	Costa Rica Guatemala Honduras	39,6 % 30,3 % 23,9 %		 ascenso, descenso y estabilización
	guasapear (wasapear)	16,3 % ±7,1 %	Costa Rica Uruguay Honduras	30,7 % 30,3 % 24,5 %		
11	guglear (googlear)	3,0 % ±1,1 %	Chile Filipinas Cuba	4,9 % 4,6 % 4,5 %		 estabilización
12	emoyi (emoji)	0,04 % ±0,04 %	El Salvador Puerto Rico Venezuela	0,2 % 0,1 % 0,1 %		 ascenso y descenso
13	paranoya (paranoia)	0,6 % ±1,3 %	España El Salvador Bolivia	6,4 % 1,1 % 1,0 %		 ascenso, descenso y estabilización

TABLA 2. Análisis diatópico y temporal. Tabla de elaboración propia. Los datos surgen del análisis detallado en la metodología

la Academia, haciendo referencia a la influencia de la aplicación Zoom en la popularización de este término (Pérez, 2020).

6 Discusión

Se han estudiado ciertas voces neológicas del OdP con el objeto de clarificar su funcionamiento y su cometido. Los términos analizados han mostrado que, a pesar de una gran variabilidad en su frecuencia absoluta de uso, todos se extienden a lo largo del territorio hispanohablante. Aunque algunos países han quedado excluidos en unos pocos pares de variantes, en nin-

gún caso se observaron agrupaciones regionales, por lo que podríamos hablar de términos de alcance pan-hispánico. Asimismo, en general, el uso mayoritario desde el punto de vista global se reflejó también en el ámbito local.

Por otra parte, los análisis temporales han puesto de manifiesto no solo la existencia de distintos estadios evolutivos para cada voz, sino también la clara influencia de los fenómenos sociales en la variación lingüística. Incluso si ya estaban establecidos en uso, los picos de *webinario* y *mutear* reflejan su aumento de uso por las prácticas que impuso la pandemia en 2020, así como el de *guasap* indica la popularización de la aplicación. Con la excepción de *turismólogo/turistólogo*, en ningún

caso hubo —al menos a escala global— una inversión en las preferencias de uso entre variantes, lo que parece indicar cierta inflexibilidad en el cambio de voces establecidas. Quizás es justamente la baja frecuencia general del par referido al turismo lo que ha permitido su inversión en el uso.

De modo similar, tal como plantea Sánchez Lancis (2022), los extranjerismos tienden a estar más arraigados que las versiones castellanizadas propuestas; sea por prestigio, moda o especialización, su uso determina la preferencia y el diccionario puede acabar incluyéndolos y marcándolos en cursiva (pp. 144-145). Cuando el extranjerismo posee una grafía o fonética que facilita su adaptación sin grandes cambios, puede dar lugar a conservar grafías no patrimoniales a final de palabra (p. 145), como en *guasap/wasap* (cabe notar que la RAE las considera adaptaciones válidas). Se ha visto, empero, que las conjugaciones verbales sí favorecen las grafías tradicionales (*espoilear, ranquear*), por lo que quizás —aunque aún minoritarias— pueda haber una tendencia hacia ellas. El uso preponderante de la *k* en el caso *rankear/ranquear* no es suficiente para concluir que sea un mecanismo de uso actual, como sí podría manifestar la preferencia por el prefijo *bi-* sobre el *di-*. Sería interesante analizar más ejemplos similares.

Aunque posiblemente todas las palabras satisfagan las características a las que se refería Freixa (2002) para la incorporación de estos términos al diccionario (§ 2) y, aunque la mayoría fueran calificados por la RAE como aceptables, la autora notaba que las voces socialmente institucionalizadas no necesariamente son finalmente incorporadas al diccionario. De las palabras del OdP analizadas, solo una está en el DLE (*ludificación*), una en el DPD (*ranquín*) y dos en el DA (*eficientizar, eficientar*). Todas ellas respetan la ortografía española y, puesto que están recogidas en diferentes diccionarios, plantean dudas respecto a los criterios empleados para determinar su inclusión y a qué consecuencias pueden derivarse de este hecho en cuanto a la frecuencia o valoración sociolingüística del uso. Aunque el DPD y el DA son parte de la norma prescrita para el español y a la vez la configuran, no tienen la misma llegada al público ni el mismo prestigio que el *Diccionario de la lengua española*. De hecho, una de las dificultades que se le presenta al hablante al consultar una palabra es tener que hacerlo en todos los diccionarios por separado, pues no hay una búsqueda centralizada y, por defecto, se busca en el DLE. Lo más sencillo sigue siendo *guglear* el término en cuestión, agregando «RAE» a la búsqueda, o, alternativamente, enviar la consulta a la RAE (o a la FundéuRAE, fundación que también responde consultas).

Las palabras fueron incorporadas al OdP, según explica la RAE, por haber generado dudas entre los hablantes. Tratándose de términos de extensión panhispánica, cabría pensar que estas preguntas han tenido procedencias variadas. Lamentablemente, las consultas en Twitter sobre estos usos no son lo suficientemente

numerosas como para llevar a cabo un análisis estadístico fiable de cada consulta (como es el caso de las consultas generales [Pardo, 2022]) —cerca de la mitad de los casos estudiados no alcanza los cien tuits y retuits con ubicación válida—, pero un análisis cualitativo de los hilos de mensajes puede servir como introspección a las diferentes actitudes y opiniones de los hablantes.

Si bien la decisión en este trabajo de tomar pares de variantes de alternancia limitó la elección de los términos, permitió, al mismo tiempo, una comparación fidedigna de los usos locales a escala global. Un estudio de regionalismos puede ser comparativo si se limita a un solo país o puede realizarse a mayor escala si se desea corroborar cualitativamente la información provista en el observatorio.

A este respecto, cabe destacar que la creación del OdP responde a la actual política lingüística de la RAE, la presentación de una imagen, discurso y disposición renovada y tolerante de la institución, abierta al compromiso con el público (Paffey, 2021, p. 253). Aunque las entradas del OdP cumplen criterios de neologicidad y permiten interesantes análisis respecto a la incorporación de las voces, este no ha sufrido cambios desde 2020 y valdría evaluar qué voces fueron incorporadas al DLE sin pasar por el Observatorio. Ciertamente, los objetivos institucionales van más allá de los meramente lingüísticos, dada la compleja naturaleza sociopolítica de la gestión de la lengua y las disputas entre los diversos actores.

7 Conclusión

Este artículo ha estudiado —empleando Twitter como corpus dialectológico— la variación diatópica desde el punto de vista del país y la evolución temporal, durante la última década, de una selección de términos del OdP de la RAE, basada en la posibilidad de definir pares de variantes de alternancia. La metodología de mapeo dialectológico permitió llevar a cabo un análisis de extensión panhispánica y corroborar que todos los términos seleccionados tienen uso a escala global (incluso los preferidos muy acusadamente en ciertos países de la hispanofonía) y que, a excepción de un único caso, las variantes mayoritarias globalmente lo son también localmente. Las recomendaciones de la RAE —que en su mayoría son validaciones de las voces— se condicen con el uso popular, a excepción de los extranjerismos, que en el habla popular priman sobre las versiones españolas, incluso si, en los usos verbales, la ortografía muestra una tendencia hacia la castellanización de estos términos.

A pesar de que la RAE objete que el OdP sea un estadio intermedio en la aprobación de las voces y se desentienda de la aceptación de sus entradas, es claro que su presencia en este observatorio no es indiscriminada. Lejos de mostrar indiferencia, la Academia ha creado este novedoso espacio y ha clasificado sus

voces como adecuadas, aceptables, correctas y recomendadas, o como innecesarias, no recomendadas y desaconsejadas. Tal y como se señala explícitamente, esta información puede cambiar en el futuro, pero hoy en día son esos los juicios de valor establecidos por la autoridad de referencia de la lengua, a quien el hablan-

te común recurre mayoritariamente, no sin el esfuerzo continuado de la Academia. Los usos, sin embargo, progresarán y variarán en función de la negociación permanente entre la conservación y la innovación en el uso del idioma. ✿

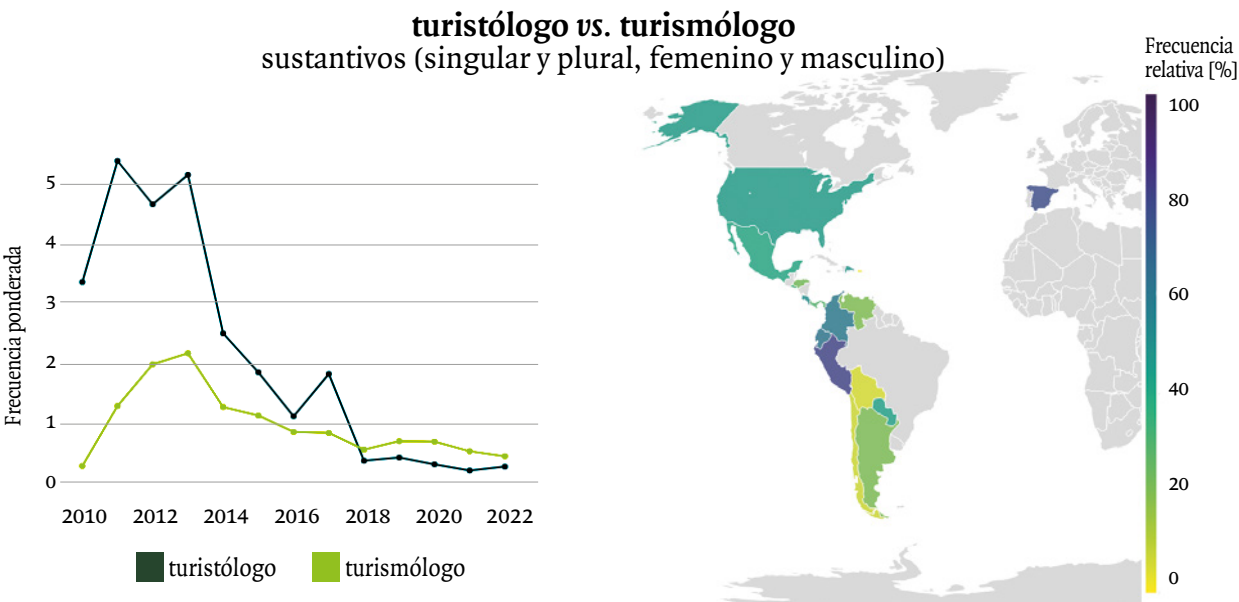
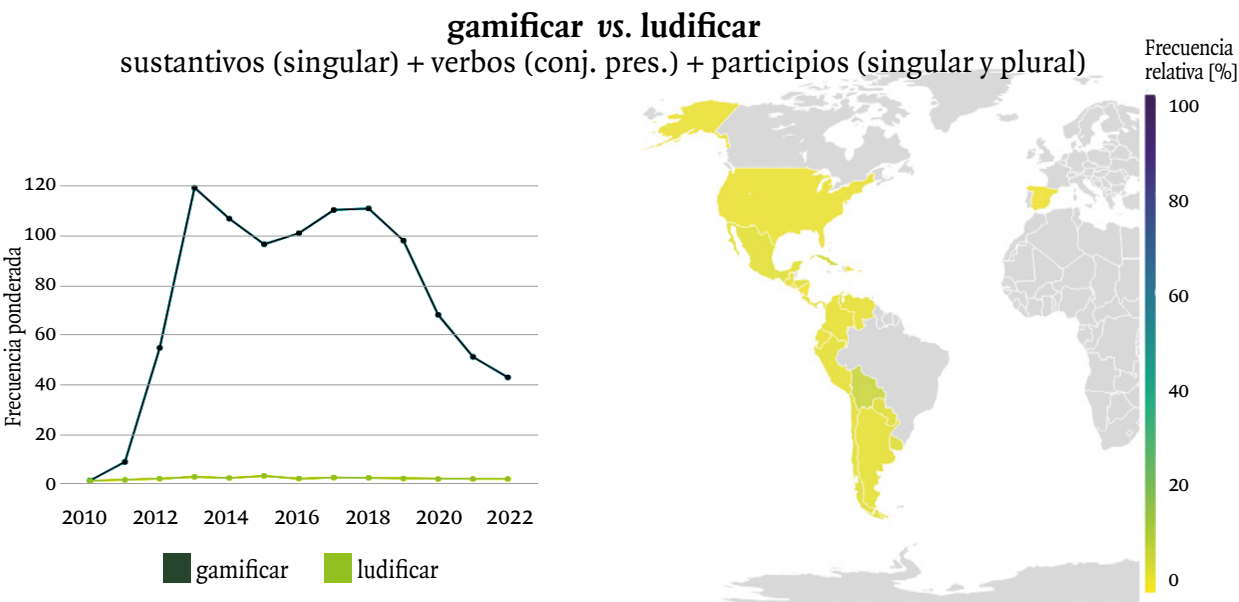
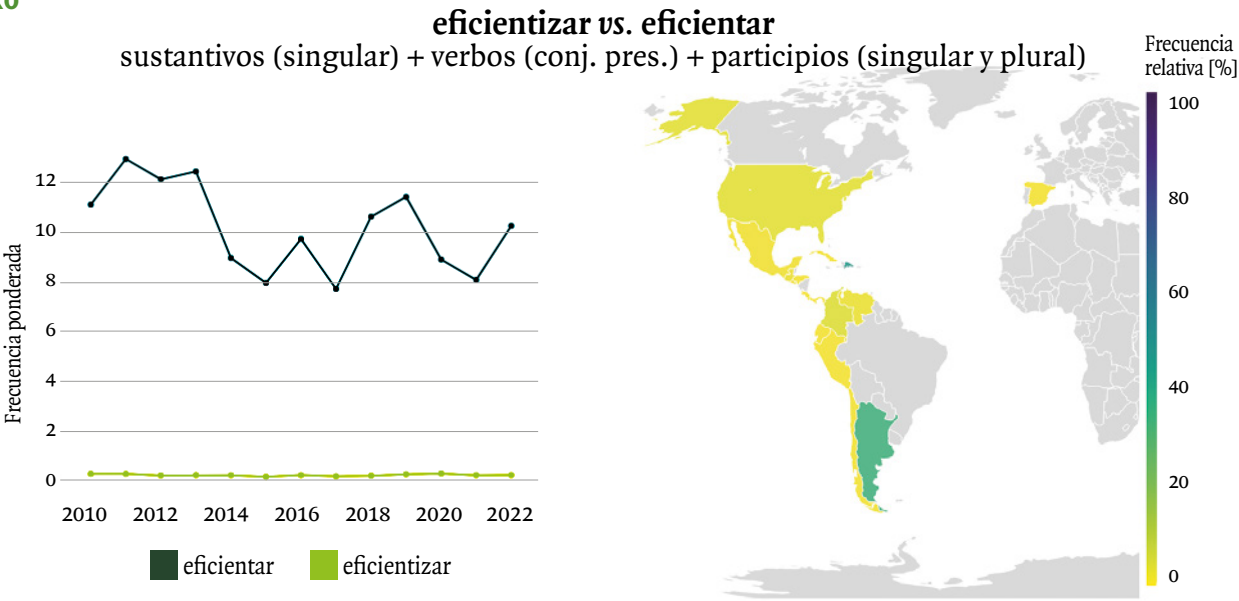
8 Bibliografía

- AMORÓS-NEGRE, Carla (2018). *La estandarización lingüística de los relativos en el mundo hispánico: Una aproximación empírica* [en línea]. Madrid: Iberoamericana; Frankfurt am Main: Vervuert. <<https://doi.org/10.31819/9783954877874>> [Consulta: 20 enero 2022].
- ARNANDIS, Rubén (@rubenarnandis) (2017). «Gracias @RAEinforma. Así lo haremos. Esperamos ver muy pronto “turismólogo/a” en las nuevas entradas del DLE!». Twitter, 25 mayo. <<https://twitter.com/rubenarnandis/status/867649499560374272>>.
- ASALE (2022). *Academias* [en línea]. <<https://www.asale.org/academias>> [Consulta: 30 enero 2022].
- ÁVILA, Raúl (2004). «¿El fin de los diccionarios diferenciales? ¿El principio de los diccionarios integrales?». *Revista de Lexicografía*, x, pp. 7-20.
- CARLA, Paola (@calapola). (2014). «@RAEinforma #RAEconsultas ¿Existe el verbo eficientar o eficientizar? Lo he escuchado mucho y me parece extraño». Twitter, 13 mayo. <<https://twitter.com/calapola/status/466288803792314369>>.
- COOK, Paul; HAN, Bo; BALDWIN, Timothy (2014). «Statistical methods for identifying local dialectal terms from GPS-tagged documents». *Dictionaries: Journal of the Dictionary Society of North America*, 35 (35), pp. 248-271.
- DALUNA MX (@dalunam) (2018). «¿“Guasap”? ¿Es en serio? ¡Quiero llorar! ¡Por favor díganme que es un sueño, una broma o que es una cuenta falsa de la RAE!». Twitter, 14 agosto. <<https://twitter.com/dalunam/status/1029349439729819648>>.
- DOYLE, Gabriel (2014). «Mapping dialectal variation by querying social media». *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics* (abril), pp. 98-106.
- DUNN, Jonathan; ADAMS, Benjamin (2020). «Mapping languages and demographics with georeferenced corpora» [en línea]. <<https://arxiv.org/abs/2004.00809>>.
- EISENSTEIN, Jacob (2017). «Identifying regional dialects in on-line social media». En: BOBERG, Charles; NERBONNE, John; WATT, Dominic (ed.). *The handbook of dialectology*. Hoboken, New Jersey: John Wiley & Sons, Inc., pp. 368-383.
- ESTRADA ARRAEZ, Ana; Benito Moreno, Carlota de (2016). «Variación en las redes sociales: datos twilectales». *Revista Internacional de Lingüística Iberoamericana*, 28(2), 77-111.
- EXYGE (@exYge) (2020). «Si escribimos ludificación nadie va a saber de qué hablamos. Por otra parte no es nuevo el uso de terminología». Twitter, 21 abril. <<https://twitter.com/exYge/status/1252661135666536461>>.
- FELIPE EL IGUANO (@jeli702) (2019). «@RAEinforma es la primera vez que escucho que se use “destrigar” en este sentido. ¿Es propia de algún país o región?». Twitter, 25 marzo. <<https://twitter.com/jeli702/status/1110185688723587074>>.
- FERNÁNDEZ-SILVA, Sabela (2022). «La variación geográfica de la neología en español y su representación lexicográfica». En: BERNAL, Elisenda; FREIXA, Judit; TORNER, Sergi (ed.). *La neología del español: Del uso al diccionario*. Madrid: Iberoamericana; Frankfurt am Main: Vervuert, pp. 259-282.
- FREIXA, Judit (2022). «El proceso neológico». En: BERNAL, Elisenda; FREIXA, Judit; TORNER, Sergi (ed.). *La neología del español: Del uso al diccionario*. Madrid: Iberoamericana; Frankfurt am Main: Vervuert, pp. 51-71.
- GARCÍA PLATERO, Juan Manuel (2022). «Panorama de la investigación de la neología en español». En: BERNAL, Elisenda; FREIXA, Judit; TORNER, Sergi (ed.). *La neología del español: Del uso al diccionario*. Madrid: Iberoamericana; Frankfurt am Main: Vervuert, pp. 17-34.
- GONÇALVES, Bruno; SÁNCHEZ, David (2014). «Crowdsourcing dialect characterization through Twitter». *PLoS ONE* [en línea], 9 (11), pp. e112074. <<https://doi.org/10.1371/journal.pone.0112074>> [Consulta: 20 enero 2022].

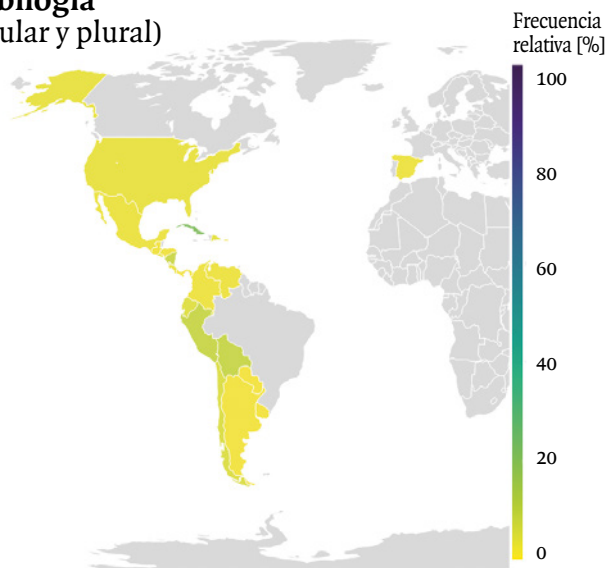
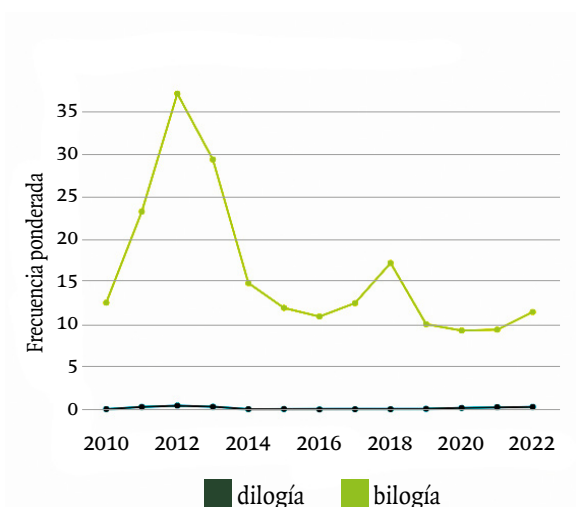
- (2016). «Learning about Spanish dialects through Twitter». *Revista Internacional de Lingüística Iberoamericana* [en línea], 28 (2), pp. 65-75. <<http://digital.casalini.it/4134616>> [Consulta: 20 enero 2022].
- GREUSSLICH, Sebastian (2015). «El pluricentrismo de la cultura lingüística hispánica: política lingüística, los estándares regionales y la cuestión de su codificación». *Lexis*, 39 (1), pp. 57-99.
- GRIEVE, Jack; MONTGOMERY, Chris; NINI, Andrea; MURAKAMI, Akira; GUO, Diansheng. (2019). «Mapping lexical dialect variation in British English using Twitter». *Frontiers in Artificial Intelligence* [en línea], 2 (11). <<https://doi.org/10.3389/frai.2019.00011>> [Consulta: 3 enero 2022].
- GUERRERO RAMOS, Gloria (2022). «La neología en los diccionarios del español». En: BERNAL, Elisenda; FREIXA, Judit; TORNER, Sergi (ed.). *La neología del español: Del uso al diccionario*. Madrid: Iberoamericana; Frankfurt am Main: Verbuert, pp. 191-209.
- HUANG, Yuan; GUO, Diansheng; KASAKOFF, Alice; GRIEVE, Jack. (2016). «Understanding US regional linguistic variation with Twitter data analysis». *Computers, Environment and Urban Systems*, 59, pp. 244-255. Birmingham: University of Birmingham.
- INSTITUTO CERVANTES (2022). *El español en el mundo 2022: Anuario del Instituto Cervantes* [en línea]. <https://cvc.cervantes.es/lengua/anuario/anuario_22/el_espanol_en_el_mundo_anuario_instituto_cervantes_2022.pdf> [Consulta: 5 junio 2022].
- JEHL, Laura Elisabeth (2010). «Machine translation for Twitter». En: *Edinburgh Research Archive* [en línea]. <<https://era.ed.ac.uk/handle/1842/5317>> [Consulta: 23 marzo 2022].
- JIMÉNEZ RÍOS, Enrique (2015). «Recorrido histórico por las razones para la admisión de voces nuevas en la lengua y en el diccionario». *Philologica Canariensis*, 21, pp. 45-80.
- KOCH, Peter; OESTERREICHER, Wulf (2007). *Lengua hablada en la Rumania: español, francés, italiano*. Traducción de Araceli López Serena. Madrid: Gredos. [Versión original: *Gesprochene Sprache in der Rumania: Französisch, Italienisch, Spanisch* (1990). Tübingen: Max Niemeyer Verlag]
- LARA, Luis Fernando (2005). «Para una reconstrucción de la idea de la lengua española: Más allá de las fronteras instituidas». *Revista Internacional de Lingüística Iberoamericana*, 6, pp. 171-187.
- LARA (@laraalonsosimon) (2019). «Todo eso llega tardísimo, me temo. “Gamificación” está tan arraigada en el ámbito de la enseñanza...». Twitter, 20 septiembre. <<https://twitter.com/laraalonsosimon/status/1175008999185571840>>.
- LONGLEY, Paul A.; ADNAN, Muhammad; LANSLEY, Guy. (2015). «The geotemporal demographics of Twitter usage». *Environment and Planning A*, 47 (2), pp. 465-484.
- LORENTE CASAFONT, Mercè (2022). «La vitalidad de las lenguas y el cambio léxico». En: BERNAL, Elisenda; FREIXA, Judit; TORNER, Sergi (ed.). *La neología del español: Del uso al diccionario*. Madrid: Iberoamericana; Frankfurt am Main: Verbuert, pp. 35-50.
- LUKAČ, Morana (2018). «Grassroots prescriptivism: An analysis of individual speakers’ efforts at maintaining the standard language ideology». *English Today*, 34 (4), 5-12.
- MALIK, Momin; LAMBA, Hemank; NAKOS, Constantine; PFEFFER, Jürgen. (2015). «Population bias in geotagged tweets». *Proceedings of the International AAAI Conference on Web and Social Media*, 9 (4), 18-27.
- NEGRETE-PORTILLO, Rafael (@_Rafael_Negrete) (2018). «Gracias. Pues empezaremos a llamarlo “ludificación” a ver [sic] conseguimos ganarle la batalla al calco». Twitter, 22 mayo. <https://twitter.com/_Rafael_Negrete/status/998895946149908480>.
- NGUYEN, Dong; DOĞRUÖZ, A. Seza; ROSÉ, Carolyn P.; JONG, Franciska de (2016). «Computational sociolinguistics: A survey». *Computational Linguistics*, 42 (3), pp. 537-593.
- NIEDZIELSKI, Nancy A.; PRESTON, Dennis R. (2000). *Folk linguistics*. Ed. a cura de W. Winter. Berlín: Nueva York: Mouton de Gruyter. (Trends in Linguistics: Studies and Monographs; 122)
- PARAKRAMA, Arjuna (1995). *De-hegemonizing language standards learning from (post)colonial Englishes about ‘English’*. Londres: Palgrave Macmillan.
- PAVALANATHAN, Umashanthi; EISENSTEIN, Jacob (2015). «Confounds and consequences in geotagged Twitter data» [en línea]. <<https://arxiv.org/abs/1506.02275>>.
- PRIETO DE LOS MOZOS, Emilio (1999). «Sobre las gramáticas normativas y las gramáticas no normativas». En: ARGENTER GIRALT, Joan Albert. *Homenatge a Jesús Tusón*. Barcelona: Empúries, pp. 245-258.
- RAE (2020^a). *Observatorio de Palabras* [en línea]. <<https://www.rae.es/portal-linguistico/observatorio-de-palabras>>. Internet Archive. <<https://web.archive.org/web/20201029041000/https://www.rae.es/portal-linguistico/observatorio-de-palabras>> [Consulta: 23 julio 2022].
- (@RAEinforma) (2020b). «Gracias por su interés. Debido a la confusión que ha generado la presencia de “elle” en el “Observatorio de palabras”...». Twitter, 30 octubre. <<https://twitter.com/RAEinforma/status/1322152976186806278>>.
- (@RAEinforma) (2021a). «Todas las palabras que aparecen actualmente en el “Observatorio” se incluyeron en él en 2020». Twitter, 15 febrero. <<https://twitter.com/RAEinforma/status/1361238500943802368>>.

- (@RAEinforma) (2021b). «El objetivo del observatorio no es recoger las palabras que están en fase de incorporación al diccionario, sino resolver dudas». Twitter, 14 mayo. <<https://twitter.com/RAEinforma/status/1393164970049875968>>.
- (2022). Observatorio de Palabras [en línea]. <<https://www.rae.es/portal-linguistico/observatorio-de-palabras>> [Consulta: 23 julio 2022].
- RAE; ASALE (2005). *Diccionario panhispánico de dudas* [en línea]. Madrid: Real Academia de la Lengua Española: Asociación de Academias de la Lengua Española: Santillana Ediciones Generales. <<https://www.rae.es/dpd/>> [Consulta: 25 julio 2022].
- (2010). *Diccionario de americanismos* [en línea]. Madrid: Real Academia de la Lengua Española. Asociación de Academias de la Lengua Española: Santillana Ediciones Generales. <<https://www.asale.org/damer/>> [Consulta: 25 julio 2022].
- (2022). *Diccionario de la lengua española* [en línea]. 23.^a ed. Madrid: Espasa [versión 23.5]. <<https://dle.rae.es>> [Consulta: 25 julio 2022].
- PARDO, Julia Elena (2022). #dudaRAE: A merging analysis of folk prescriptivism and dialectology [en línea]. Tesis de máster. Universitat Pompeu Fabra (UPF). Repositori Digital de la UPF. <<https://repositori.upf.edu/handle/10230/54016>> [Consulta: 1 julio 2022].
- PAFFEY, Darren J. (2021). «State-appointed institutions: Authority and legitimacy in the Spanish-Speaking world». En: AYRES-BENNETT, W.; BELLAMY, J. (ed.). *The Cambridge handbook of language standardization* [en línea]. Cambridge: Cambridge University Press, pp. 234-262. (Cambridge Handbooks in Language and Linguistics) <<https://doi.org/10.1017/9781108559249.009>> [Consulta: 10 marzo 2022].
- PÉREZ, Leopoldo (@leopoldo_perez7) (2020). «El uso de la herramienta del ZOOM es que, a puesto la palabra “mutear” de moda, yo la escuché hace un mes». Twitter, 26 mayo. <https://twitter.com/leopoldo_perez7/status/1265431929693900800>.
- RUIZ TINOCO, Antonio (2018). «Geocorpus del español de las redes sociales y cartografía automática». Monográficos SINOELE: IX Congreso Internacional de la Asociación de Hispanistas [en línea] (Bangkok, 2016), 17, pp. 598-608. <http://www.sinoele.org/images/Revista/17/monograficos/AAH_2016/AAH_2016_antonio_ruiz.pdf> [Consulta: 20 enero 2022].
- ROSSUM, Guido van; DRAKE, Fred L. (2009). *Python 3 Reference Manual* [en línea]. Scotts Valley, CA: CreateSpace. <<https://www.python.org/>> [Consulta: 03 enero 22].
- SÁNCHEZ LANCIS, Carlos (2022). «La neología, entre el uso y la prescripción». En: BERNAL, Elisenda; FREIXA, Judit; TORNER, Sergi (ed.). *La neología del español: Del uso al diccionario*. Madrid: Iberoamericana; Frankfurt am Main: Verbuert, pp. 129-150.
- SÁNCHEZ LOBATO, Jesús (2019). «La lengua española en busca del sobresaliente». En: MERINO, J. M.; GRIJELMO, A. (ed.). *Más de 555 millones podemos leer este libro sin traducción: La fuerza del español y cómo defenderla*. Taurus, pp. 65-79.
- SANMARTÍN, Julia (2022). «Neología y registro: sobre lo coloquial, lo formal y lo especializado en medios de comunicación escritos». En: BERNAL, Elisenda; FREIXA, Judit; TORNER, Sergi (ed.). *La neología del español: Del uso al diccionario*. Madrid: Iberoamericana; Frankfurt am Main: Verbuert, pp. 233-258.
- SENZ, Silvia; MINGUELL, Jordi; ALBERTE, Mon (2011). «Las academias de la lengua española, organismos de planificación lingüística». En: *El dardo en la Academia: Esencia y vigencia de las academias de la lengua española*, 1, pp. 371-550.
- STATISTA (2022a). Number of monthly active Twitter users worldwide from 1st quarter 2010 to 1st quarter 2019 [en línea]. <<https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/>> [Consulta: 12 marzo 2022].
- (2022b). Annual Twitter user growth rate worldwide from 2015 to 2024 [en línea]. <<https://www.statista.com/statistics/303723/twitters-annual-growth-rate-worldwide/>> [Consulta: 12 marzo 2022].
- VALLE, José del del (2007). «La RAE y el español total: ¿esfera pública o comunidad discursiva?». En: VALLE, José del (ed.). *La lengua, ¿patria común?: Ideas e ideologías del español*. Madrid: Iberoamericana; Frankfurt am Main: Verbuert, pp. 81-96.

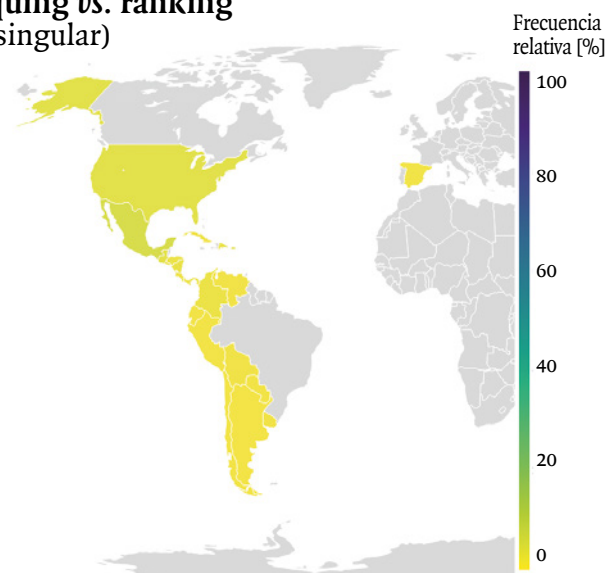
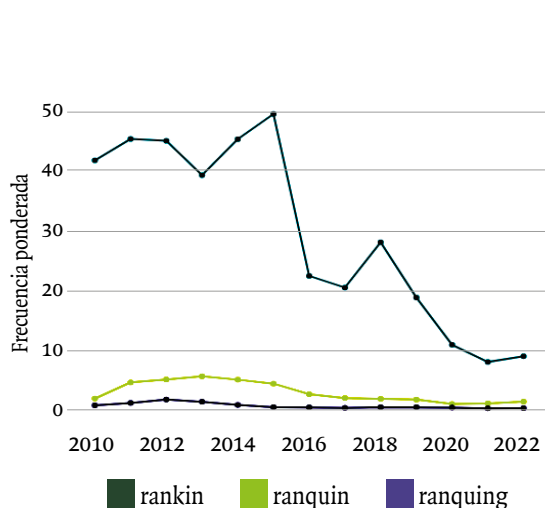
Anexo



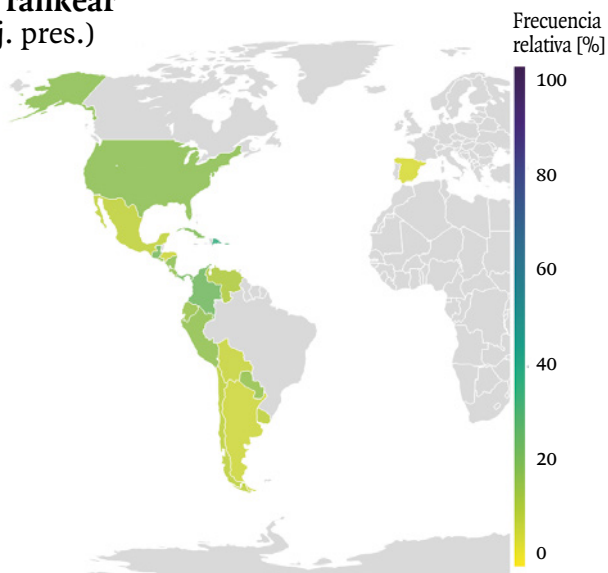
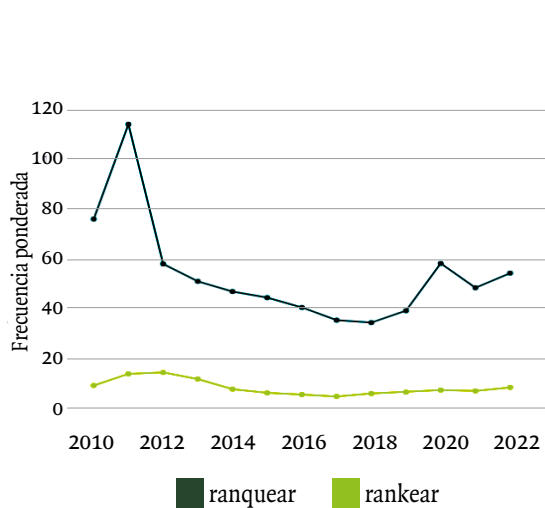
dilogía vs. biología sustantivos (singular y plural)



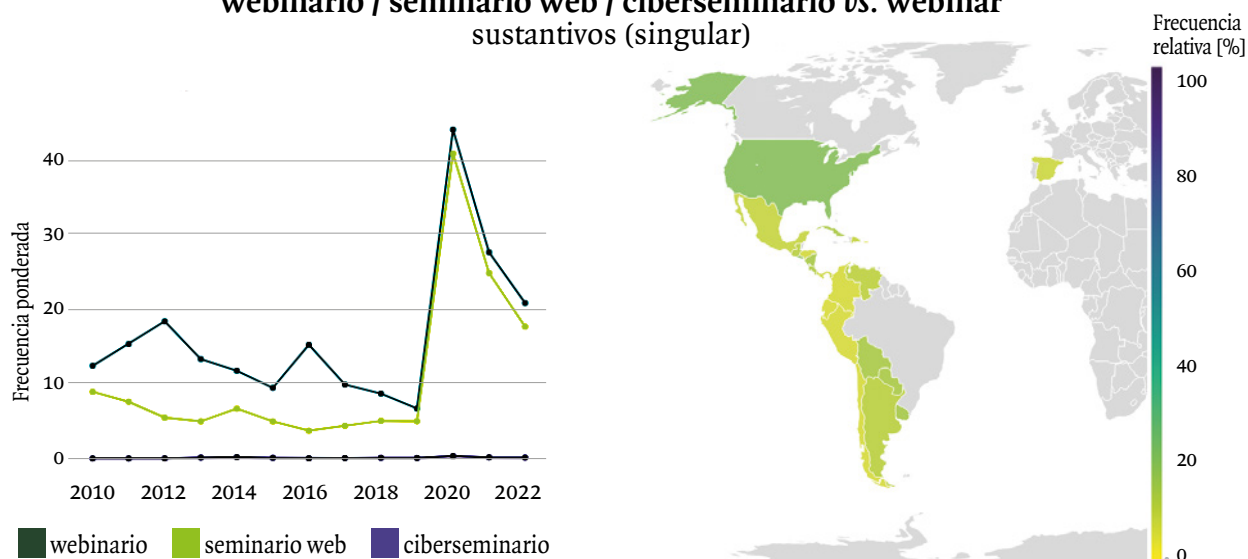
rankin/ranquin/ranquing vs. ranking sustantivos (singular)



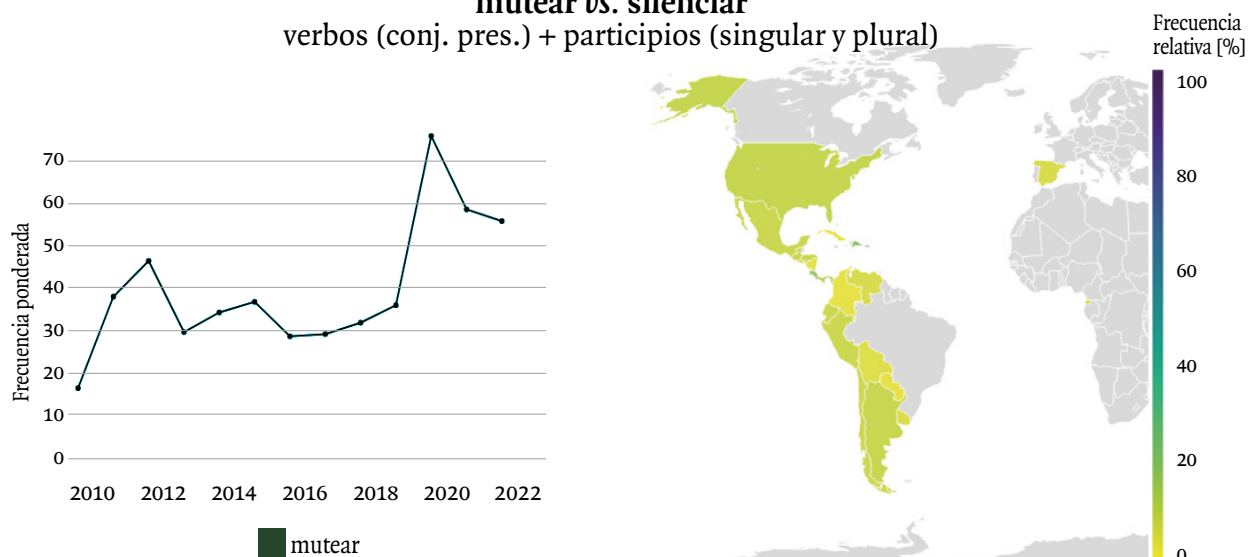
ranquear vs. rankear verbos (conj. pres.)



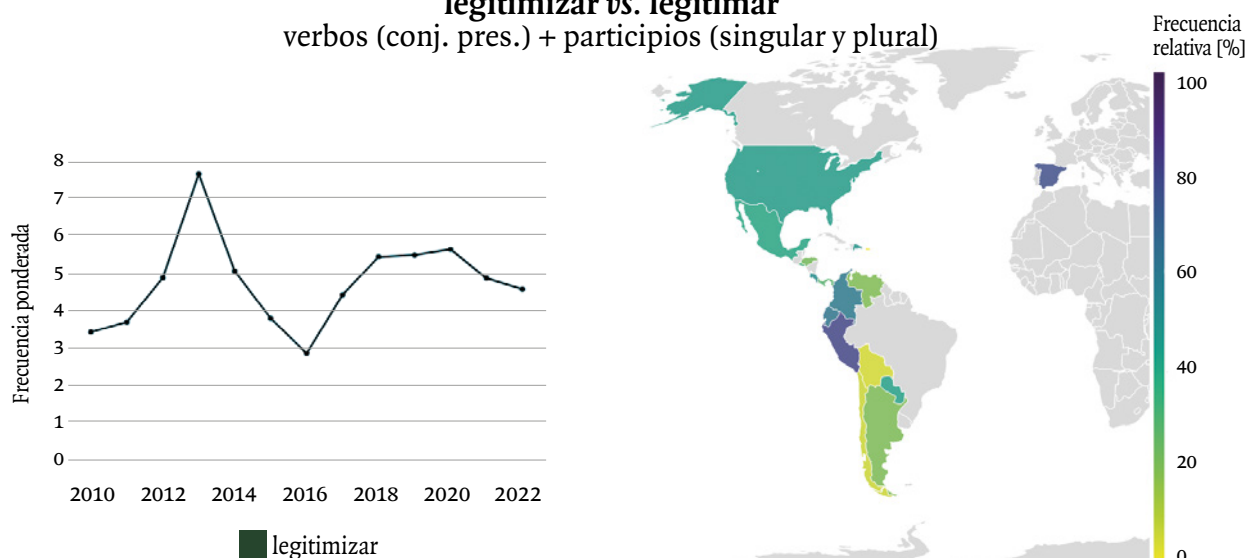
webinario / seminario web / ciberseminario vs. webinar sustantivos (singular)



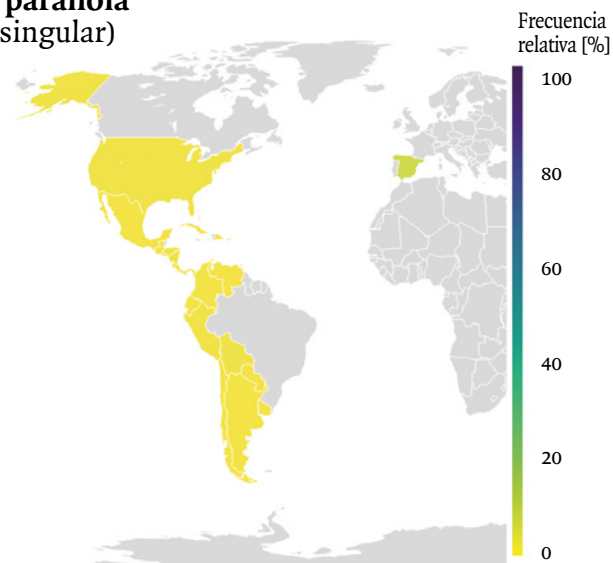
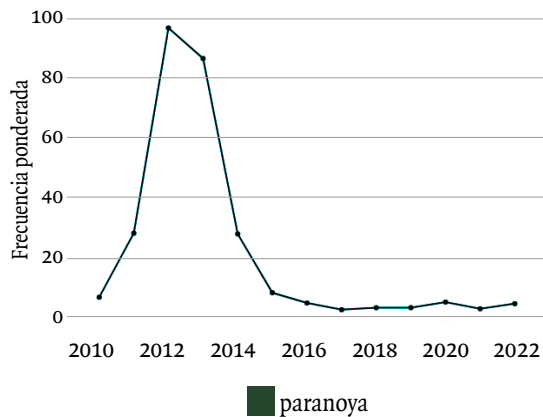
mutear vs. silenciar verbos (conj. pres.) + participios (singular y plural)



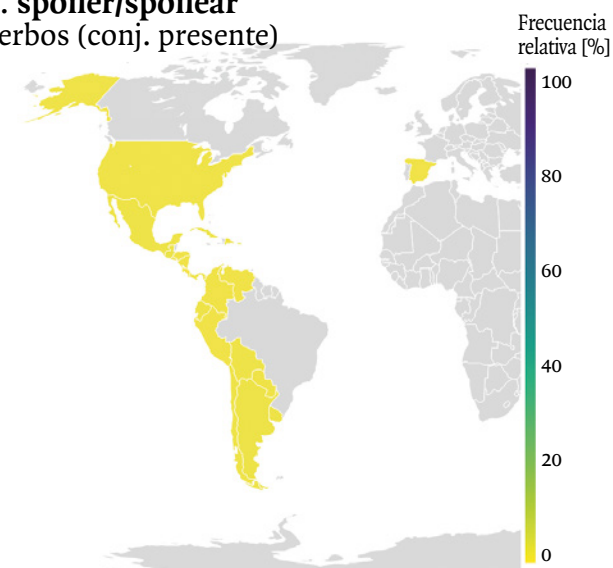
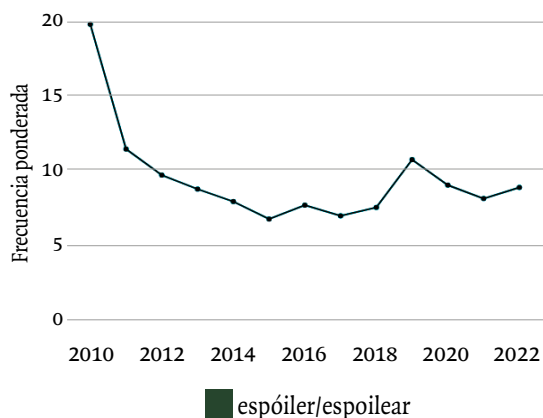
legitimizar vs. legitimar verbos (conj. pres.) + participios (singular y plural)



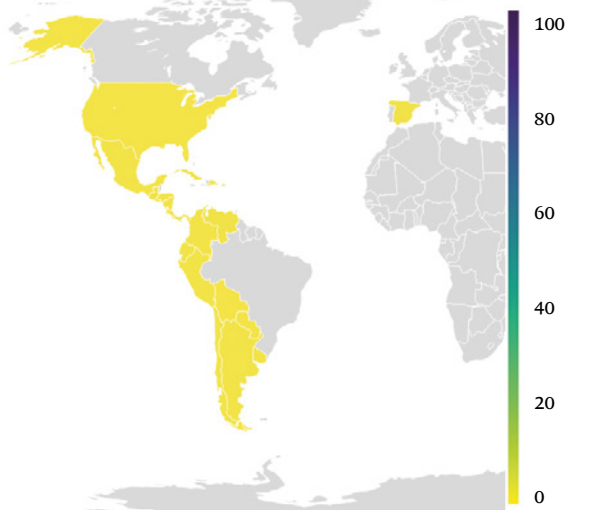
paranoya vs. paranoia
sustantivos (singular)



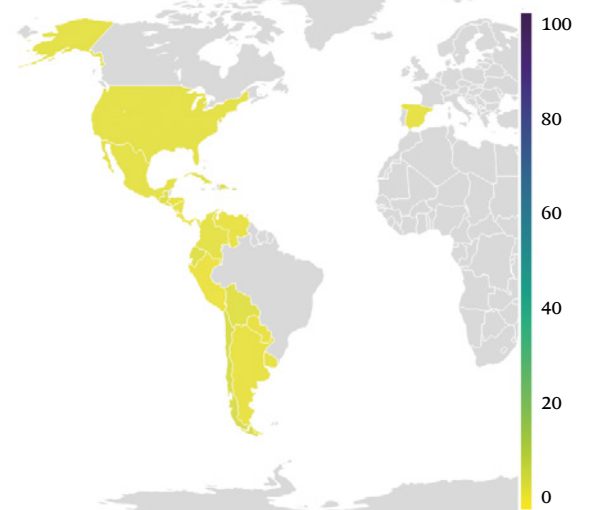
espóiler/espoilear vs. spoiler/spoilear
sustantivos (singular) + verbos (conj. presente)



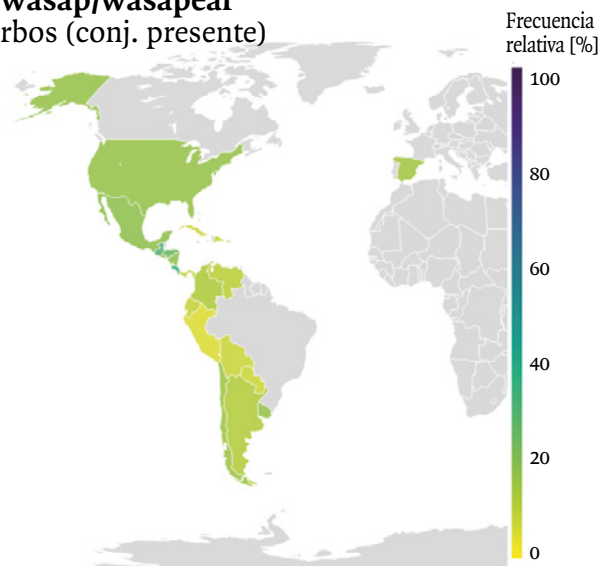
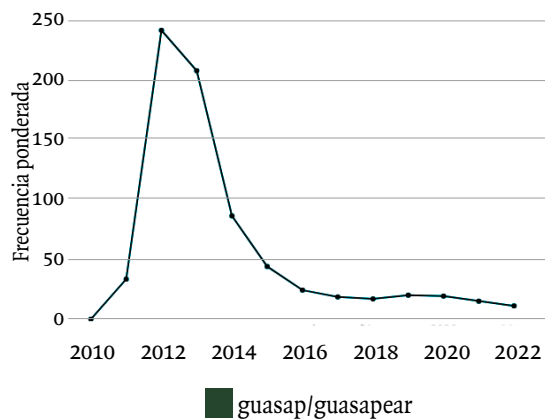
espóiler vs. spoiler
sustantivos (singular)



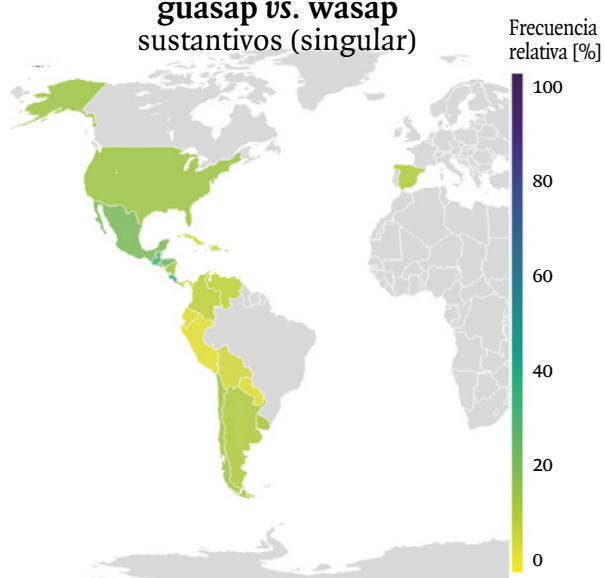
espoilear vs. spoilear
verbos (conj. presente)



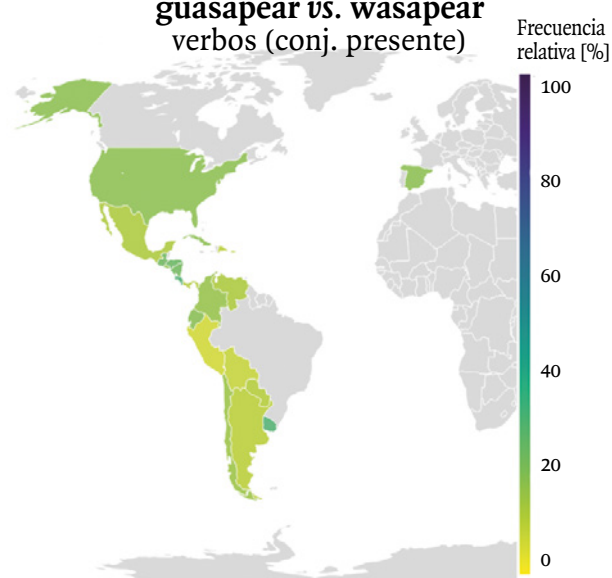
guasap/guasapear vs. wasap/wasapear
sustantivos (singular) + verbos (conj. presente)



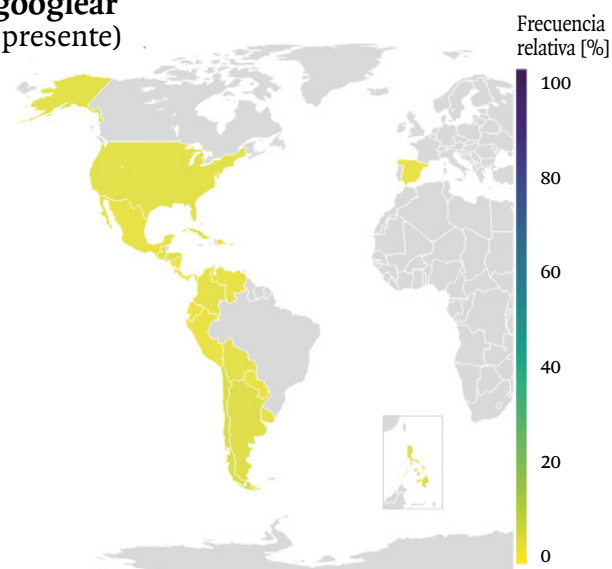
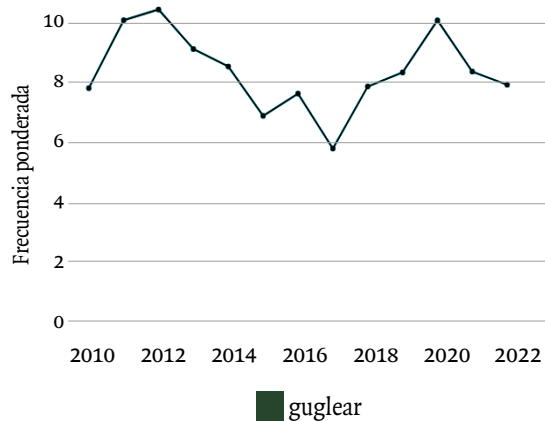
guasap vs. wasap
sustantivos (singular)

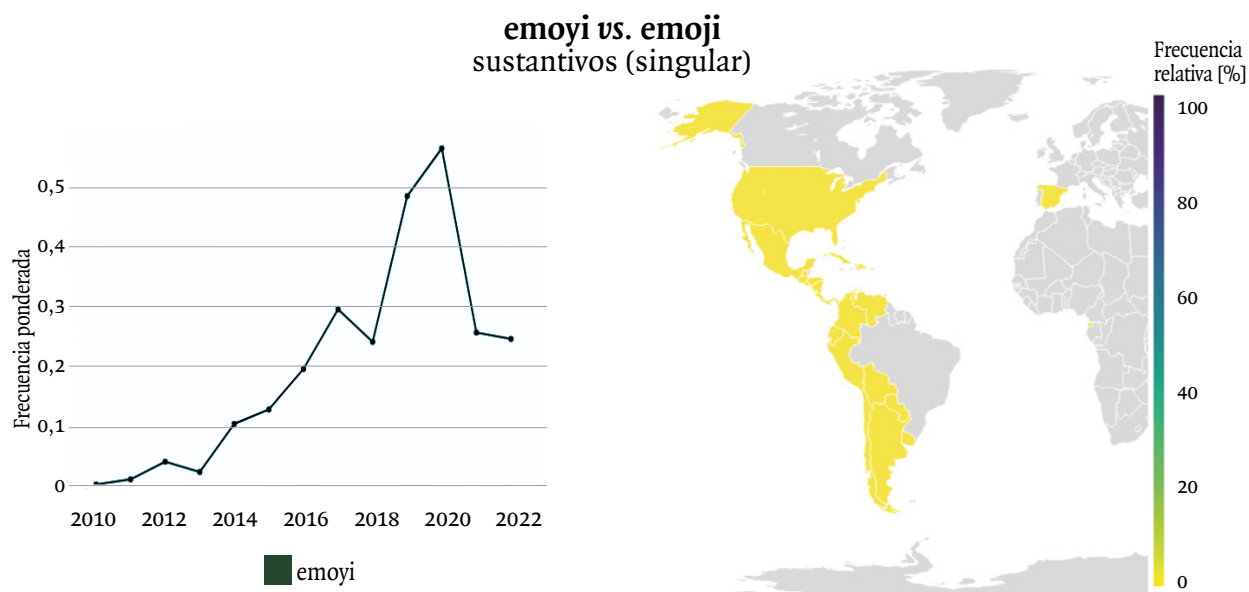


guasapear vs. wasapear
verbos (conj. presente)



guglear vs. googlear
verbos (conj. presente)





Descripción y clasificación: las unidades terminológicas metafóricas en los mercados de flores cortadas

SANGZHU ZHAXI

Universitat Pompeu Fabra

ORCID: 0000-0003-2039-8784

sangzhu.zhaxi@upf.edu

Sangzhu Zhaxi és florista

d'estil naturalista i professor d'arranjaments florals. Màster de Traducció entre Llengües Globals: Xinès-Espanyol de la Universitat Pompeu Fabra i, actualment, doctorand en el Grup IULATERM (Lèxic i Tecnologia) de la mateixa universitat, té una tesi en curs sobre la variació en l'àmbit de les flors i l'estandardització terminològica per al comerç internacional, en què aprofundeix els coneixements en el sector de les flors tallades, la terminologia i la lingüística aplicada en general. També hi investiga com descriure i harmonitzar els usos terminològics coexistents en el sector, en el marc de la teoria comunicativa de la terminologia (TCT).



Resumen

El desarrollo de la industria floral ha llegado a alcanzar una considerable notoriedad como sector emergente en el ámbito económico mundial. A diferencia de España, en los mercados florales chinos el uso de términos metafóricos es mucho más frecuente. Basamos nuestro estudio en los términos florales extraídos de las investigaciones in situ entre cuatro mercados de flores y no solamente describimos el uso metafórico de los términos florales, sino que además los clasificamos desde el punto de vista lingüístico para analizar mejor las necesidades de estandarización terminológica del sector y en el mercado internacional.

PALABRAS CLAVE: terminología metafórica; normalización terminológica para el comercio internacional; flor cortada; chino; español

Resum

Descripció i classificació: les unitats terminològiques metafòriques en els mercats de flors tallades

El desenvolupament de la indústria floral ha assolit una notorietat considerable com a sector emergent en l'àmbit econòmic mundial. A diferència d'Espanya, en els mercats florals xinesos l'ús de termes metafòrics és molt més freqüent. Basem el nostre estudi en els termes florals extrets de les investigacions in situ de quatre mercats de flors i no només descrivim l'ús metafòric dels termes florals, sinó que també els classifiquem des del punt de vista lingüístic per conèixer millor les necessitats d'estandardització terminològica del sector i en el mercat internacional.

PARAULES CLAU: terminologia metafòrica; normalització terminològica per al comerç internacional; flor tallada; xinès; espanyol

Abstract

Description and classification: Metaphorical terminology units in cut flower markets

The cut flower industry has achieved considerable prominence as an emerging sector in the global economic arena. As opposed to the situation in Spain, in Chinese flower markets the use of metaphorical terms is quite common. On the basis of the floral terms extracted from on-site investigations at four flower markets, this paper not only describes the metaphorical use of floral terms but also classifies them from the linguistic standpoint to better understand the need for terminological standardization in this sector and in the international market.

KEYWORDS: metaphorical terminology; terminology standardization for international trade; cut flowers; Chinese; Spanish

TERMINÀLIA 27 (2023): 51-67 · DOI: 10.2436/20.2503.01.188

Data de recepció: 14/12/2022. Data d'acceptació: 14/12/2022

ISSN: 2013-6692 (impresa); 2013-6706 (electrònica) · <http://terminalia.iec.cat>

1 Introducción

La metáfora es un recurso privilegiado del lenguaje y la cognición humana. El trabajo fundamental de Lakoff y Johnson (2008) *Metaphors we live by* nos abrió una nueva perspectiva sobre la metáfora. Fundamentalmente, destacaron el carácter constitutivo de la metáfora en el sistema cognitivo, los valores de este recurso lingüístico cognitivo como instrumento del pensamiento y, además, su omnipresencia en todos los dominios de la actividad humana.

Tradicionalmente, la terminología había abordado la metáfora desde el punto de vista denominativo. No obstante, según estudios más recientes sobre la terminología, la lingüística cognitiva y las ciencias del discurso (Rossi, 2017 y Suárez Gallo, 2019, entre otros), se demuestra que la metáfora va más allá de ser un ornato o una figura de estilo. En las áreas especializadas, además de ser un recurso importante para formar y consolidar el léxico especializado, también lo es para transmitir conocimiento especializado en escenarios comunicativos de niveles de especialización diferentes. Por lo tanto, la metáfora es «un recurso cognitivo, lingüístico y comunicativo» (Suárez Gallo, 2019, p. 173).

Debido a las nuevas concepciones epistemológicas sobre la actividad científica, la influencia de la historia y la sociología de la ciencia, y los estudios sobre la retórica científica, según afirma Ciapuscio (2011, p.89), en los últimos años la actitud hacia la metáfora ha cambiado de manera sustantiva, y este cambio coincide con las nuevas reflexiones de la lingüística cognitiva. El interés por la metáfora en los discursos especializados actuales deriva, principalmente, de los diferentes escenarios comunicativos de la ciencia y la tecnología: el valor heurístico de las metáforas en la investigación y la solución de problemas en ciencia, tanto para conceptualizar y dar cuenta de relaciones entre fenómenos (Black, 1962, p. 45) como para focalizar percepciones de manera técnicamente legítima y heurísticamente fértil (Prelli, 1989). Como indica Ciapuscio (2011, p. 91), la metáfora evoca dominios experienciales cotidianos, de modo que constituye un recurso comunicativo efectivo para explicar y presentar contenidos científicos a audiencias de especializaciones distintas.

No obstante, en otros ámbitos especializados como las actividades comerciales y los mercados, el estudio de la terminología, en general, y de la terminología metafórica, en particular, ha sido poco tratado. Según Cabré (2010, p. 2), «los científicos, los tecnólogos o los profesionales de cualquier ámbito requieren términos para representar y expresar sus conocimientos, para informar, transferir o comprar y vender sus productos». No solo en China, sino también en otros países, las flores suelen tener múltiples denominaciones para satisfacer necesidades diferentes o para referirse a ellas en distintos ámbitos, como el científico-botánico, la alimentación, la medicina, la floristería, la jardinería, etc.

La terminología de las flores es un ámbito de especialización con mucha variación. Las causas de este volumen de variación denominativa pueden ser diversas: la terminología referida a las flores y las plantas procedente de ámbitos científicos (botánica, medicina, farmacología) y de ámbitos profesionales (jardinería, agricultura, industria y comercio floral, decoración floral, cosmética) puede mostrar diferencias. Estas diferencias pueden basarse en la orientación y la especificidad de cada ámbito, y se reflejan tanto en las denominaciones como en los conceptos, por lo que se recomienda tratar la terminología de las flores sin mezclar la terminología propia de cada ámbito relacionado.

En este artículo nos centramos en el sector de las flores cortadas,¹ sector emergente que cuenta cada día con más neologismos, mediante recursos propios de cada lengua y también a través de préstamos. Se trata de un sector tan activo, internacional, innovador y poco regulado, que incluye el uso de ciertas denominaciones que pueden causar confusión y obstáculos en la comunicación. Las investigaciones realizadas y las reflexiones al respecto nos han dejado ver que la metáfora es un punto interesante para guiar las inflexiones acerca de la diferencia entre la lengua china y la española, así como de la función que desempeña la metáfora en los discursos especializados de comunicación de carácter comercial.

Como se trata de un tema poco tratado, el objetivo del texto consiste en: a) presentar un acercamiento a la casuística de la terminología metafórica del sector de las flores cortadas en chino y en español, b) comparar el uso de la terminología metafórica en los mercados florales entre España y China y c) proponer una clasificación de las unidades terminológicas (UT) desde la mirada de un florista que toma como referente teórico los fundamentos de la teoría comunicativa de terminología —a partir de ahora, TCT— (Cabré, 1999).

Partimos de los fundamentos teóricos de la TCT y de la teoría de la metáfora conceptual (Lakoff y Johnson, 1995). Analizamos los términos recogidos en mercados florales chinos y españoles, según las investigaciones *in situ* realizadas: en primer lugar, presentamos el panorama del sector floral y sus términos y, en segundo lugar, los clasificamos desde tres dimensiones distintas: la estructura, los campos metafóricos y la transparencia de las unidades terminológicas metafóricas (UTM).

2 Marco teórico

Los principios propuestos por la TCT en cuanto al valor terminológico y los principios de variación y adecuación constituyen los puntos fundamentales para la base teórica del análisis que aquí presentamos. La propuesta de la TCT parte del principio de que el marco en el que pueden describirse y explicarse adecuada-

mente las unidades terminológicas, concebidas como «unidades de naturaleza interdisciplinaria (cognitivas, lingüísticas y sociales), debe ser amplio y flexible» (Cabré, 2009, p. 10). Según Cabré (1993, p. 96), un término tiene un componente cognitivo porque representa conocimiento; también, el término es una unidad del lenguaje natural, por lo tanto, posee un componente lingüístico. Además, la transmisión de determinada información refleja el carácter comunicativo de las unidades terminológicas. Desde esta perspectiva, la TCT asume que «la comunicación especializada no mantiene un estatus completamente aparte del que mantiene la comunicación general; y el conocimiento especializado no es ni uniforme ni está totalmente separado del conocimiento general en todas las situaciones de comunicación» (Cabré, 2002).

Una unidad terminológica, entendida como una unidad léxica cuya estructura responde a una unidad léxica de origen o producto de la lexicalización de un sintagma, posee un significado específico en el ámbito en el que se asocia y es necesaria en la estructura conceptual del dominio del que forma parte (Cabré y Estopà, 2005, p. 77). Para ser considerada una UT las condiciones que debe cumplir una unidad léxica son: la especialidad semántica y la necesidad en la estructura conceptual del campo del conocimiento correspondiente. El valor terminológico que explica la TCT ilustra la idea de que los términos no están aislados del lenguaje natural ni de las unidades léxicas, sino que son las unidades léxicas que activan el valor en función de su uso en un contexto. Además, según Cabré (1993, p. 133), el valor terminológico se establece por el lugar que ocupa una unidad léxica en la estructuración conceptual de una materia, y cada ámbito se puede estructurar de maneras diferentes. También, «un concepto puede participar en más de una estructura con el mismo o diferente valor. Por lo tanto, los términos no pertenecen a un ámbito, sino que son usados en un ámbito con un valor singularmente específico» (Suárez Gallo, 2019, p. 144, citando a Cabré, 1999). Como consecuencia de esto, en un área determinada, se pueden encontrar varias unidades léxicas que poseen un mismo valor (sinónimos, variación denominativa) y denominaciones que esconden cambios de concepto (variación conceptual, polisemia). Como un recurso del lenguaje natural y como consecuencia de la economía cognitiva, usamos los conceptos de otros ámbitos más familiares para acercarnos a aquellos más abstractos a fin de conocer el mundo;² y la metáfora puede ser un ejemplo idóneo para explicar este proceso.

Lakoff y Johnson (1995, p. 39) asumieron que «las metáforas impregnan la vida cotidiana, no solo el lenguaje, sino también el pensamiento y la acción». Estos teóricos entienden la metáfora como un recurso básico en cualquier proceso de carácter cognitivo porque casi siempre se conceptualiza en términos metafóricos lo que se siente, se cree o se piensa (Suárez Gallo,

2019, p. 67). Como expresiones lingüísticas, las metáforas son posibles porque forman parte del sistema conceptual de una persona (Lakoff y Johnson, 1995, p. 42). Más adelante, la teoría integrada de la metáfora primaria, que fue elaborada por Lakoff y Johnson en 1999, y que reúne otras teorías (la de la metáfora conceptual, las metáforas primarias de Grady [1997], la neuronal de Narayanan), propuso la idea de que las metáforas son inherentes al ser humano, que tienen carácter universal. Suárez Gallo (2019, p. 77) también enfatiza la idea que presentan Lakoff y Johnson: «las metáforas no solamente se manifiestan en el lenguaje humano, sino que también existen en el arte, en los gestos, entre otras dimensiones de la vida cotidiana de los seres humanos».

En cuanto al tratamiento de la metáfora desde la terminología, según los postulados de la TCT, el estudio de las expresiones lingüísticas que vehiculan metáforas en el discurso especializado del genoma humano quizá sea el referente más cercano. Además, en los estudios de Oliveira (2005 y 2009), la metáfora desempeña un papel de sustitución, una función de suplente, que permite favorecer la comprensión de uno o varios aspectos del concepto que estructura:

En determinadas circunstancias, el instrumento metafórico puede alcanzar un nivel de comprensión mucho más profundo que el término culto, por su transparencia y familiaridad, llama la atención en que también puede ocurrir que aquellas metáforas que resultan más expresivas y más accesibles, puede que no sean lo suficientemente ricas para reflejar toda la complejidad conceptual del término culto.

Suárez Gallo (2019, p. 166) señala que, desde esta perspectiva, «la metáfora se concibe como un metalenguaje, que permite esclarecer y redescubrir un concepto poco común por medio de otro que resulta más familiar, pues en muchas ocasiones los términos cultos no son lo suficientemente ilustrativos ni palpables». Sabemos que la terminología representa el concepto, y si los seres humanos utilizamos un concepto más familiar para esclarecer uno menos común utilizando la metáfora, es imposible negar la existencia de la terminología metafórica.

En cuanto a los trabajos sobre la metáfora especializada, debemos destacar el esfuerzo que ha hecho Rita Temmerman, quien es conocida por su trabajo sobre la metáfora en la terminología y la comunicación técnica. Ha estudiado cómo las metáforas se utilizan en la comunicación técnica y científica, y cómo influyen en la comprensión y la interpretación de los conceptos especializados (Temmerman, 2021, p. 223). Además, Blackwell (2006) ha escrito sobre la metáfora en la programación y el diseño de *software* y argumenta que las metáforas son importantes en la interacción humano-computadora y que pueden mejorar la comprensión y el uso del *software*.

La metáfora no se puede reducir a un teorema, ni tampoco a una permuta de palabras (Suárez Gallo, 2011, p. 1), porque tiene que ver con el pensamiento y es un asunto de cognición. En este artículo nos concentramos en los términos metafóricos en los mercados de flores cortadas, a fin de dar el primer paso para conocer el papel que desempeña la metáfora en la terminología del campo comercial, especialmente, en el sector de las flores cortadas.

3 Metodología

El presente trabajo es parte del proyecto doctoral, en curso en la Universitat Pompeu Fabra, bajo la dirección de Mercè Lorente, que propone un estudio descriptivo de la variación de la terminología metafórica de las flores cortadas en español y en chino, y que incluye un análisis sobre las UT metafóricas del sector y un estudio aplicado sobre la estandarización comercial. Las fuentes del trabajo son recursos terminográficos diversos (para el comercio, la floración profesional, la traducción o la estandarización científica). A falta de un recurso terminológico completo y confiable del ámbito y para conocer mejor las necesidades del mercado, realizamos una investigación *in situ* para coleccionar los términos realmente usados en cuatro mercados chinos y españoles y construimos una base de datos terminológica.

3.1 Fuentes para la investigación *in situ*

En primer lugar, la fuente de los conceptos de los que partimos proviene de dos colecciones, *BASICS plant flashcards. Cut flowers*, vol. I y II (flores cortadas I y 2), de Bloom's.

BASICS plant flashcards incluye la mayoría de los términos de las flores cortadas desde un punto de vista industrial y se considera el material más importante para la formación de floristas cuando empiezan su carrera para conocer los términos y los usos básicos de las flores.

Los mercados seleccionados para realizar las investigaciones *in situ* son el mercado de flores cortadas Shenghua Honglin, en Beijing (norte de China); el mercado de flores cortadas Wanfu, en Chengdu (sur de China); el Mercat de Flor i Planta Ornamental de Catalunya, y Mercabarna Flor (estos dos últimos en la provincia de Barcelona). En la tesis doctoral está previsto ampliar el estudio con informaciones de mercados de flores cortadas de Colombia.

3.2 Nivel de especialidad de los términos florales

Las UT, con toda su variación, que hemos recopilado tienen su origen en: 1) etiquetas de precios, 2) listas de productos y 3) facturas. Todas son situaciones comunicativas sumamente naturales.

Cabe destacar que en los mercados florales es donde se mezclan interlocutores distintos: vendedores de productos florales - floristas (especialista-especialista), vendedores de productos florales - público (especialista - público entendido) y los floristas y vendedores, que pueden tener conocimientos diferentes, algunos más profesionales que otros (especialista-semiespecialista).

Como se trata de una parte del estudio doctoral, además de la recogida de términos usados por los profesionales, a fin de realizar estudios comparativos desde perspectivas interdisciplinarias, es necesario una base de datos que permita clasificar adecuadamente las diferentes entradas, como se resume en la tabla 1.

Como se ilustra en la tabla 1, las ideas coinciden con la propuesta de la TCT, lo cual demuestra la naturaleza interdisciplinaria de las UT. A fin de realizar una descripción de las UT como unidades lingüísticas, cognitivas y sociales, la base de datos terminológica (BDT) es amplia y flexible; incluye los diferentes escenarios de estandarización terminológica del sector y las opiniones y los problemas encontrados por los equipos interdisciplinarios en terminología (expertos, lingüistas, traductores).

A partir de las investigaciones realizadas en los cuatro mercados chinos y españoles durante nueve meses, hemos terminado la fase de recogida de 144 flores cortadas comunes que existen en los mercados, entre los 200 términos facilitados por *BASICS plant flashcards*. Cuando comparamos los términos encontrados en ambos países, comprobamos que el uso de denominaciones metafóricas es más frecuente en chino que en español. El trabajo actual consiste en un estudio descriptivo y en la clasificación de estas flores cortadas, con la intención de que en el futuro sirva así que en el futuro servirá para conocer mejor este ámbito emergente poco tratado y disponer de ideas para realizar estandarización terminológica de comercios internacionales.

3.3 Modelo de clasificación

Aunque las flores no configuran un tema que esté lejos de nuestra vida cotidiana, los estudios dedicados a su terminología son escasos. Para que se entienda mejor el uso de las UTM, partimos de tres dimensiones para clasificarlas. En primer lugar, analizamos la estructura de las UTM en chino y en español. Entre ellas, hay unidades monoléxicas y unidades poliléxicas, y entre estas existen unidades completamente metafóricas y parcialmente metafóricas. Por lo tanto, en este apartado las clasificamos en varios subgrupos para que se tenga una idea clara de la forma y la estructura de estas unidades metafóricas. En segundo lugar, las agrupamos según los campos metafóricos que utilizó en su tesis doctoral Suárez Gallo (2019, p. 217) para saber mejor qué conceptos originales adaptan los humanos para conocer las flores. Según Suárez Gallo (2019, p. 209) el

Descripción y clasificación: las unidades terminológicas metafóricas en los mercados de flores cortadas
Sangzhu Zhaxi

Concepto	Fuente
Fotografía	BASICS plant flashcards (versión inglesa) (BLOOM's, 2017)
Denominación latina	BASICS plant flashcards (versión inglesa) (BLOOM's, 2017)
Denominaciones inglesas	BASICS plant flashcards (versión inglesa) (BLOOM's, 2017)
Denominaciones chinas	BASICS plant flashcards (versión china) (BLOOM's, 2017, Trad. Wen, 2018)
idem (transcripción)	BASICS plant flashcards (versión china) (BLOOM's, 2017, Trad. Wen, 2018)
Término est. Botánica (chino)	Flora Reipublicae Popularis Sinicae (Wu et al., 2004)
idem (transcripción)	Flora Reipublicae Popularis Sinicae (Wu et al., 2004)
idem (comentario)	Flora Reipublicae Popularis Sinicae (Wu et al., 2004)
Término estandarizado	Terminology Online ⁴
idem (transcripción)	Terminology Online
idem (área temática)	Terminology Online
idem (comentario)	Terminology Online
Entrada en chino	Diccionario Xinhua
idem (transcripción)	Diccionario Xinhua
Entrada en español estándar	Diccionario de la lengua española de la Real Academia Española (RAE, 2014)
Entrada en español coloquial	Nuevo diccionario de americanismos (Nuevos diccionarios de colombianismos) (Günter Haensch et al., 1993)
Entrada en chino Jardinería	1000 of garden plants illustrations (Jiang, 2013) Flores y jardinería (Gong et.al., 2012) Diccionario de flores (Li, 2017) Diccionario de 1000 flores (Wang, 2018) Illustrated color landscape plants (Cao, 2012)
Entrada en español Jardinería	Diccionario especializado
Entrada medicina tradicional china en chino	Diccionario nacional de medicina tradicional china (Li, Hu, 2015) Medicinal Plant Images Database (Chen et al., 2012)
Mercado de flores cortadas Shenghua Honglin en Beijing (chino)	Materiales
idem (transcripción)	Materiales
Mercado de flores cortadas Wanfu en Chengdu (chino)	Materiales
idem (transcripción)	Materiales
Mercat de Flor i Planta Ornamental de Catalunya	Materiales
Mercabarna Flor	Materiales
Mercado de Bogotá	Materiales

Concepto	Fuente
Puesto de flores domésticas Beijing	Encuesta 1
Puesto importador Beijing	Encuesta 1
Puesto de flores domésticas Chengdu	Encuesta 1
Puesto importador Chengdu	Encuesta 1
Puesto importador Mercabarna	Encuesta 1
Puesto de flores domésticas Mercabarna	Encuesta 1
Puesto importador Mercat de Flor i Planta Ornamental de Catalunya	Encuesta 1
Puesto de flores domésticas Mercat de Flor i Planta Ornamental de Catalunya	Encuesta 1
Puesto importador Bogotá	Encuesta 1
Puesto exportador Bogotá	Encuesta 1
Profesional: profesor Beijing	Encuesta 2
Profesional: profesor Chengdu	Encuesta 2
Profesional: profesor España	Encuesta 2
Profesional: profesor Colombia	Encuesta 2
Profesional: dueño de granja floral	Encuesta 4
Profesional: traductor China	Encuesta 3
Profesional: traductor España	Encuesta 3
Profesional: traductor Colombia	Encuesta 3
Profesional: terminólogo chino de Terminology Online	Encuesta 4

TABLA 1. Estructura de la base de datos terminológica (BDT) de terminología floral chino-español

concepto *campo metafórico* se entiende como «el ámbito semántico o temático en el que se agrupan y asocian las UT creadas por metáfora, a partir de las relaciones de analogía que se establecen entre unas y otras, que parten siempre de la existencia de un rasgo semántico compartido». Por último, nos concentramos en la familiaridad de este proceso de cognición a través de la transparencia de las UTM, ya que entre estas unidades existen algunas que son fáciles de entender y algunas que no lo son.

4 Análisis

4.1 Factores que condicionan la variación denominativa en el mercado

Antes de analizar las UT de las flores cortadas del sector, hay que tener en cuenta que las flores cortadas en los mercados son conceptos, pero que también son productos. Sabemos que el precio consiste en un indicador sumamente importante para distinguir los pro-

ductos comerciales entre sí y que en el sector de las flores cortadas pasa lo mismo. En algunas situaciones comunicativas específicas, como en la subasta de flores, las mismas rosas diferenciadas solamente por la longitud del tallo son conceptos/productos distintos entre la oferta de flores cortadas, y las denominaciones que se utilizan para representarlos deben mostrar esta diferencia; por ejemplo: «Rosa Freedom 40 cm» y «Rosa Freedom 50 cm».

Entre los factores que afectan el precio de las flores se pueden citar: la temporada (las de fuera de temporada siempre son más caras; por eso, por ejemplo, los tulipanes en verano resultan más caros que en primavera), el lugar de producción (las rosas holandesas y las japonesas siempre son más caras), la longitud del tallo (las flores más largas tienen precios más altos, ya que se pueden utilizar en los arreglos grandes y son más difíciles de cultivar y transportar) y la calidad de la flor (las flores del nivel A siempre valen más que las de los niveles B o C). Aunque la longitud del tallo es un indicador de la calidad floral, el diámetro del tallo, el estado de los pétalos y las hojas o la madurez son otros

criterios importantes. Otro factor fundamental en el establecimiento de precios en la subasta (que pueden llegar a variar cada día) son las festividades señaladas en los diversos países; por ejemplo, las rosas en San Valentín o en Sant Jordi y los claveles en el Día de la Madre son más caros que en otras fechas. En consecuencia, la afectación de todos estos factores generan necesidades de denominación diferenciada.

De hecho, en muchos puestos de los mercados florales hay abundantes rosas de colores variados, de longitudes de tallo distintas e incluso de orígenes diferentes, y los vendedores deben distinguirlas muy bien, ya que su precio es bastante diferente. En los intercambios comerciales nacionales e internacionales también se precisa distinguir exactamente las flores de precio distinto, ya que con frecuencia las flores cuentan con denominaciones generales tales como *rosa holandesa*, *clavel colombiano*, *rosa ramificada de 50 cm*. Cabe destacar que, en los mercados, esta inclinación orientada de precio es bastante común y los términos reales muestran diferencias respecto de las unidades terminológicas que encontramos en los libros de botánica y de jardinería, e incluso respecto de las denominaciones que se registran en los libros de arreglos florales y en los manuales de floristas. A continuación, se presentan algunos ejemplos:

En mercados españoles:

- Denominación latina binominal completa: *Carthamus tinctorius*.
- Denominación latina binominal completa con nombre de la variedad detallada: *Ornithogalum thyrsoides* 'mount Fuji' blanco.
- Denominación latina con acortamiento: *Anthurium* (*Anthurium andraeanum*), *Virginiana* (*Physostegia virginiana*), *Antir.* (*Antirrhinum majus*).
- Denominación con descripción: *Amaranthus verde colgante*, *clavel colombiano*, *flor de cera nacional*.
- Denominación española/inglesa/catalana: *jacinto*, *espuela de caballero* / *sosmos*, *poppy* / *blauet*, *muguet*.

En mercados chinos:

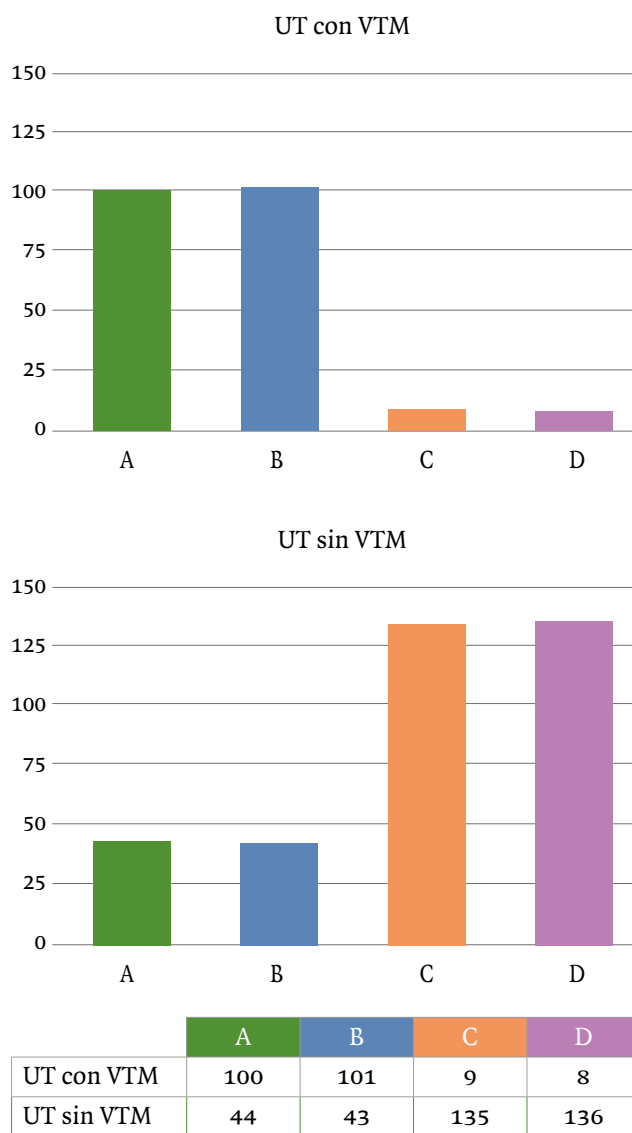
- Denominación de especie simple: 洋牡丹 [yáng mǔ dān] (ranúnculos).
- Denominación de especie detallada con descripción: 黄色蝴蝶洋牡丹 [huáng sè hú dié yáng mǔ dān] (ranúnculos mariposa amarillos), B级波浪洋桔梗 浅紫 [bī jí bō làng yáng jú gēng qiǎn zǐ] (*Lisianthus* de honda del nivel B color lila).
- Acortamiento: 复古康 [fù gǔ kāng], que es un acortamiento de 复古色康乃馨 [fù gǔ sè kāng nǎi xīn] (clavel de color vintage).
- Acortamiento con descripción: A级曼塔 [āi jí màn tǎ], acortamiento de (rosa 'manta' del nivel A).

Los términos encontrados en los mercados, tanto en China como en España, cuentan con diferentes formas para necesidades y situaciones comunicativas distintas. Con lo que se ha presentado, ya tenemos una idea sobre las unidades de estructuras distintas en este sector de especialidad. Además, cabe subrayar que el

acortamiento es altamente frecuente en el ámbito y que estas unidades abreviadas pueden tener origen en las denominaciones metafóricas, por lo que hace falta que realicemos un análisis orientado hacia esa cuestión.

4.2 Las UTM de flores cortadas

Según las investigaciones que hemos hecho en los mercados, entre las 200 flores cortadas que nos brinda BASICS plant flashcards, hay 144 flores comunes en los cuatro mercados (Figura 1). En cuanto a los términos metafóricos, de una manera general, el uso de la metáfora en los mercados chinos es mucho más frecuente que en España. Podemos comparar el porcentaje del uso metafórico en la figura 2.



A – Mercado de Flores Cortadas Shenghua Honglin, en Beijing
B – Mercado de Flores Cortadas Wanfu, en Chengdu
C – Mercat de Flor i Planta Ornamental de Catalunya
D – Mercabarna Flor

FIGURA 1. Cantidad de UT con variación terminológica metafórica (VTM) y UT sin VTM

Como muestra la figura 2, en cuanto a las 144 flores cortadas con las que hemos realizado las investigaciones, en los dos mercados de flores cortadas chinos hay 100/101 UT que cuentan con VTM; en cambio, en los dos mercados de España, solamente se encuentran ocho o nueve casos. Para tener la idea más concreta, podemos comparar el porcentaje del uso de unidades metafóricas: en los dos mercados chinos son un 69,4 % y un 70,1 %, respectivamente, y en los dos mercados españoles solo alcanzan el 6,25 % y el 5,55 %. En contraposición con el uso abundante de metáforas en los mercados florales chinos, la causa de la utilización restrictiva de los términos metafóricos en España parece estar relacionada con el uso del latín. Podemos observar los datos de la figura 3, donde se presentan las unidades afectadas por la denominación latina en los dos mercados españoles.

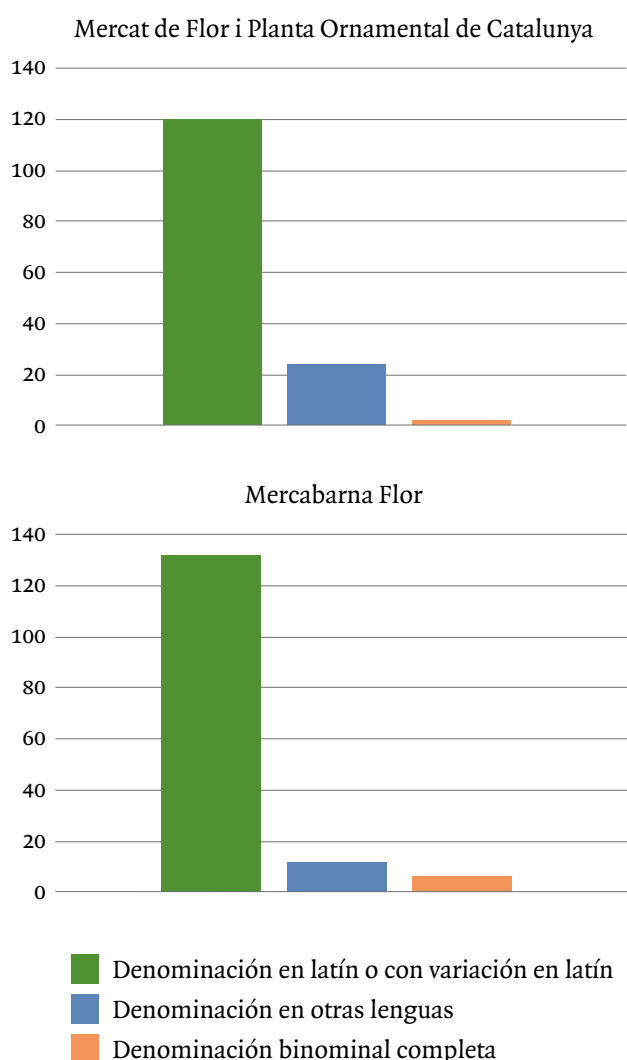


FIGURA 2. Cantidad de UT afectadas por el latín, en otras lenguas y binominales en los mercados de flores en España

A diferencia del caso chino, en los mercados españoles el latín se usa ampliamente y, en consecuencia, el uso de la metáfora es muy limitado. Aunque las denominaciones latinas son más estables y facilitan las transferencias comerciales a escala mundial, los profesionales no son capaces de conceptualizar y recordar estas flores con sus nombres en latín con facilidad. En cambio, los profesionales chinos suelen cambiar unos caracteres de estas denominaciones para perfilar conceptos parecidos. Por ejemplo, el término chino original de la flor *Gomphrena globosa* es 千日紅 [qiān rì hóng] ('rojo de mil días'), pero los profesionales lo expanden a nombres como 千日白 [qiān rì bái] ('blanco de mil días') y 千日紫 [qiān rì zǐ] ('lila de mil días') para las variedades en color blanco y lila.

4.3 Análisis de la estructura de las unidades metafóricas en los mercados de flores cortadas

Para empezar el análisis de las UTM, hay que tener en cuenta que, aunque nos basamos en los conceptos ofrecidos por BASICS plant flashcards, en los mercados actuales, las flores que llevan más historia de su cultivo siempre cuentan con denominaciones más concretas.⁵ Es decir, aparte del nombre del género y el epíteto o nombre específico, en los mercados españoles algunas unidades incluyen nombres de la variedad, tal como la parte del préstamo inglés de *Ageratum houstonianum* 'Velvet white'. Además, en los mercados españoles el uso del latín es sumamente popular, e incluso las denominaciones en latín pueden ser metafóricas. En tal sentido, clasificamos las unidades metafóricas en dos grupos generales para analizar la estructura de sintagma que es más apropiada para nuestro caso: unidades poliléxicas completamente metafóricas y unidades poliléxicas parcialmente metafóricas. Según las investigaciones *in situ* realizadas, en el corpus sobre el que se basa este análisis, todas las partes metafóricas de las unidades terminológicas poliléxicas son substantivos y no observamos ningún caso en que la parte metafórica sean adjetivos o verbos.

4.3.1 Las unidades terminológicas completamente metafóricas

En cuanto a las UT completamente metafóricas, se documentaron unidades monoléxicas y unidades poliléxicas tanto en chino como en español. Ahora bien, en comparación con el chino, donde existen más ejemplos, en español, las unidades completamente metafóricas son muy pocas. A continuación, se presentan los casos de unidades completamente metafóricas en dos subapartados: las unidades monoléxicas (4.3.1.1) y las unidades poliléxicas (4.3.1.2).

4.3.1.1 Las unidades monoléxicas completamente metafóricas

En los mercados españoles, se encontraron términos usados en español, fruto del acortamiento de denominaciones latinas: *Aster*, que proviene originalmente de la palabra latina que significa ‘estrella’ (Coromines y Pascual, 1980, p. 383), ya que esta flor se parece a las estrellas; o la flor nombrada *campánula*, que tiene origen en la denominación latina *Campanula medium* y cuyas florecitas son parecidas a las campanitas.

En el caso español, solamente hay dos variaciones terminológicas monoléxicas que son completamente metafóricas: *conejito* (*Antirrhinum majus*) y *espuela* (*Consolida ajacis*), utilizada como acortamiento de *espuela de caballero*.

En cuanto a los términos monoléxicos chinos, hay dos unidades metafóricas:

掌[zhǎng] es la denominación real usada para la flor *Anthurium andraeanum* en los mercados de flores cortadas en China. La «掌» [zhǎng] tiene forma de corazón o de mano y su nombre significa ‘palma de la mano’ en lengua china.

La flor 虞美人 [yú měi rén] *Papaver nudicaule*, que se traduce como ‘Consorte Yu’,⁷ es una planta que tiene el tallo muy delgado y largo. Con el viento, su flor tan grande, llamativa y ligera se ve como si estuviera bailando Consorte Yu, en honor de la heroína de la ópera tradicional china *Farewell my concubine*. Aunque esta unidad cuenta con tres caracteres, solo posee un morfema, que es el nombre del personaje.

Dado que los sistemas morfológicos de las lenguas española y china son diferentes, adaptamos la definición de la palabra de Packard (2000, p. 16) «theoretical word», que en realidad se corresponde con la *cí*, mínima unidad significativa o morfema.⁶ Por lo tanto, cualquier *cí* china, que solamente tiene un morfema, puede encontrarse dentro del grupo de unidades monoléxicas.

4.3.1.2 Las unidades poliléxicas completamente metafóricas

En cuanto a las unidades poliléxicas completamente metafóricas en español, solamente identificamos

un patrón único, el de [substantivo + preposición (de) + substantivo], correspondiente a los ejemplos de *ave del paraíso* (*Strelitzia reginae*) y la *espuela de caballero* (*Consolida ajacis*).

En el mercado chino, existe solamente un ejemplo parecido al patrón recién presentado: [substantivo + palabra auxiliar «之» [zhī] + substantivo]. Aunque la palabra auxiliar «之» [zhī] no es una preposición, tiene la misma función de indicar relación subordinada. Ejemplo: 伯利恒之星 [bó lì héng zhī xīng] (*Ornithogalum saundersiae*), cuya traducción literal es ‘estrella de Belén’.

Igualmente, según las unidades documentadas en los mercados, las unidades metafóricas chinas suelen omitir la palabra auxiliar para indicar relación interior. Alternativamente, la relación se describe a través de la estructura modificadora que es más común y dominante. Según Packard (2000, p. 22), «las palabras chinas también pueden caracterizarse en función del tipo de relación de modificación que se da entre los morfemas o, en otras palabras, ‘qué modifica a qué’». Hay dos estructuras modificadoras, que son: a) la estructura paralela modificadora y b) la estructura jerárquica modificadora, formada por un morfema central y otro secundario que lo modifica. Cabe señalar que precisamente las unidades metafóricas de las flores cortadas en chino siguen este segundo patrón jerárquico, como observamos en los ejemplos siguientes:

La flor *Amaranthus caudatus*, cuyo nombre chino es 情人泪 [qíng rén lèi], es un ejemplo idóneo para explicar este fenómeno. Como muestra la figura 4, podemos conseguir que la palabra «情人泪» [qíng rén lèi] (‘lágrima de los amantes’) cuente con dos partes: la de «情人» [qíng rén] y la de «泪» [lèi]. Obviamente que el morfema central es la *cí* «泪» [lèi], pero la parte modificadora «情人» [qíng rén] también consiste en una estructura jerárquica modificadora del mismo tipo, donde el morfema referido a la persona es el central:

Igualmente, la flor 袋鼠爪 [dài shǔ zhǎo] *Anigozanthos* también pertenece a este tipo: el «爪» [zhǎo] es el morfema central, que significa ‘garra’, y «袋鼠» [dài shǔ], ‘canguro’. La palabra «袋鼠» [dài shǔ] significa ‘rata con bolsa’, que también presenta una relación modificadora. Por lo tanto, la metáfora conecta

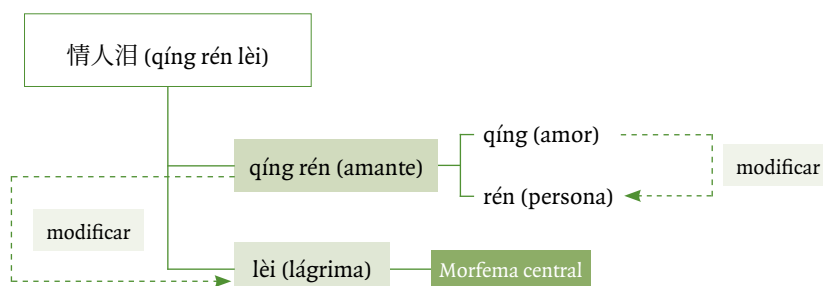


FIGURA 3. Relación modificadora de la palabra china 情人泪 [qíng rén lèi]

la flor *Anigozanthos* (dominio meta) con la garra de la rata con bolsa (dominio fuente).

La denominación china 玉米佛塔 [yù mǐ fó tǎ] de la flor *Banksia coccinea* también llama la atención, ya que esta unidad incluye dos metáforas en la misma denominación.

En realidad, *Banksia* es un grupo de flores sudafricanas que se venden ampliamente en el resto de los mercados. Además de la *Banksia coccinea*, que es más pequeña, existen otras más grandes, como la *Banksia* *genus* o la *Banksia marginata*. Aparte de su nombre adaptado fonéticamente a partir de *Banksia* «班克木» [bān kè mù], estas flores duraderas también cuentan con un nombre metafórico: «佛塔» [fó tǎ], que significa ‘pagoda budista’, ya que se reconocen fácilmente por los puntos característicos de la flor y por los conos que dan fruto, que corresponden a la forma de las pagodas en los templos budistas, con un sentido muy afincado en la cultura china.

En comparación con otras *Banksias*, la flor *Banksia coccinea* es pequeña y corta. Para distinguirla de las otras, le dieron el nombre 玉米佛塔 [yù mǐ fó tǎ], que significa ‘pagoda budista de maíz’, donde el morfema «玉米» [yù mǐ] es el modificador para limitar el ámbito del elemento modificado.

4.3.2 Las unidades poliléxicas parcialmente metafóricas

Además de las unidades completamente metafóricas, también se localizaron una gran cantidad de unidades poliléxicas o sintagmas, donde solo uno de los formantes es una metáfora. Sin embargo, debido a su carácter internacional, la terminología de este sector emergente referida a las flores cortadas y a los follajes y ramas cortadas, coincidente con otros ámbitos científicos (botánica, medicina o farmacología) y profesionales (jardinería, agricultura, industria y comercio floral, decoración floral o cosmética) puede mostrar diferencias. Estas diferencias pueden basarse en la orientación y la especificidad de cada ámbito, y se reflejan tanto en las denominaciones distintas como en las distinciones conceptuales. Al mismo tiempo, para el caso de los países europeos y de habla inglesa, las denominaciones pueden contener unidades de varias lenguas al mismo tiempo, ya que conviven situaciones y necesidades comunicativas diversas:

- La aplicación del latín y la lengua franca del inglés para facilitar las transferencias comerciales internacionales.
- La preferencia de la denominación del inglés, francés o español para las flores recientemente cultivadas de cada granja floral.
- El idioma hablado en los mercados de flores cortadas en cada país.

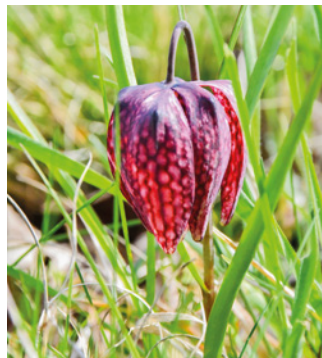
Al contrario, en los mercados chinos este fenómeno no se ve tan complicado porque el sistema de caracteres chinos es distinto del alfabeto latino. En consecuencia, traducen literalmente o fonéticamente

las denominaciones e incluso las conceptualizan con nombres más familiares y fáciles de recordar para sus hablantes. Por consiguiente, clasificamos estas UT metafóricas en los siguientes subgrupos:

4.3.2.1 La metáfora en una posición interior de la unidad poliléxica

Entre las unidades metafóricas de este tipo, la primera estructura léxica detectada sigue este patrón: [substantivo (nombre del género) + substantivo metafórico (nombre específico) + substantivo (nombre de la variedad)], como en el ejemplo *Fritillaria meleagris alba*. Sabemos que el primer formante, *Fritillaria*, es el nombre del género y que el segundo, que es el nombre específico, proviene de la zoología, ya que se refiere al género propio de los pavos.

Font: MichelleHewise, Creative Commons



Font: Frank Schultburg, Creative Commons

FIGURA 4. La flor *Fritillaria meleagris* y el animal *Meleagris*

Como muestran las fotos de la figura 4, nos damos cuenta de que la flor tiene la forma de campana colgante, con un característico dibujo a cuadros en tonos marrón-rojizo, púrpura, blanco o gris, que se asemeja a la pluma del pavo. En Mercabarna Flor, se encontraron con facilidad términos tales como *Fritillaria meleagris* y *Fritillaria meleagris* ‘alba’ durante febrero y marzo. Observamos que, aunque la palabra *meleagris* proviene del latín, también traslada una metáfora clara.

Un segundo patrón identificado en español es el de [substantivo (nombre del género) + substantivo (nombre específico) + substantivo metafórico (nombre de la variedad) + adjetivo], como en el ejemplo *Euphorbia martinii* ‘ojo de rana’ negro.

Font: David Stang, Wikimedia Commons



FIGURA 5. *Euphorbia martinii* ‘ojo de rana’ negro

Además de la parte de nombre específico, que es una metáfora, en el formante correspondiente a la variedad también se encuentra la huella metafórica. En el sector es común que las flores procedentes de la misma variedad tengan diferentes colores y los colores son también factores bastante importantes para realizar las decoraciones y los arreglos florales. En consecuencia, se encontraron fácilmente términos que llevan adjetivos de colores, como *Euphorbia martinii* ‘ojo de rana’ negro. Entre los formantes del sintagma, ‘ojo de rana’ se considera una metáfora porque existe un «ojito» en cada cáliz y el color de la planta es verde (figura 5). Por eso, es razonable que le dieran este nombre metafórico.

En cuanto al chino, se encontraron también ejemplos como 澳洲腊梅 [ào zhōu là méi], que corresponde a la flor *Chamelauicum uncinatum*. A pesar de la diferencia de la formación de palabras entre el chino y el español, podemos aprovechar el español para explicar este término chino: hay tres formantes en esta UT poliléxica: 澳洲 [ào zhōu] significa ‘australiano’, «腊» [là] significa ‘cera’ y «梅» [méi] se refiere a *Chimonanthus*.⁸ Así pues, la traducción al español de 澳洲腊梅 es ‘Chimonanthus de cera australiano’. Si observamos la estructura de esta palabra, también existen dos relaciones modificadoras, que son: «腊» [là] modifica «梅» [méi] y «澳洲» [ào zhōu] modifica 腊梅 [là méi] y la cí «腊» [là] es la metáfora por la que percibimos que esta flor tiene la textura de cera.

4.3.2.2 La metáfora en la posición final de las unidades poliléxicas

Entre las unidades metafóricas de este tipo, el patrón léxico más habitual es [substantivo + substantivo metafórico]. De hecho, una variante léxica de la flor *Chamelauicum uncinatum* usada en los mercados españoles, flor de cera, contiene una metáfora. En este término solo cera es metáfora. Además de flor de cera, en el sector existen muchos ejemplos del tipo mencionado, pero la mayoría de ellos simplemente tienen el nom-

bre de la variedad de tipo metafórico y estos nombres de variedad son préstamos de otros idiomas, como el inglés o el francés. A su vez, los nombres de variedad metafóricos suelen demostrar la semejanza de color o de forma entre las flores y los conceptos fuente, como vemos en los siguientes ejemplos:

- *Grevillea Spiderman* ‘yellow amit’, en la que la forma del pétalo se parece a las arañas y tiene color amarillo.
- Rosa ‘cappuccino’, rosa ‘café latte’, rosa ‘café con leche’ y *Dahlia* ‘café au lait’. Estas cuatro flores tienen color de crema y beige, de aquí la referencia al color del café con leche.
- Clavel colombiano ‘mostaza’, de color amarillo oscuro como la mostaza.

En cuanto al chino, tomamos como ejemplo 红苋丝绒 [hóng xiàn sī róng], que denomina la flor *Amaranthus hypochondriacus*. «红苋» [hóng xiàn] se traduce como ‘amaranto rojo’, pero la parte «丝绒» [sī róng] es una metáfora para describir la textura de esa flor, que significa ‘terciopelo’.

4.3.2.3. La metáfora en la posición inicial de las unidades poliléxicas

En los mercados españoles, documentamos *Delphinium guardian lavender*, cuyo primer formante latín es una metáfora, que proviene del griego delphinion, porque, según Pseudo-Dioscórides, sus hojas basales, hendidas y alargadas tienen forma de delfín («*Delphinium*», consulta: 2022).

La mayoría de las unidades poliléxicas metafóricas encontradas en China pertenecen a este grupo. Ponemos los ejemplos más representativos en la tabla 2 para ilustrar esta idea.

Como muestra la tabla, se presentan cuatro flores que poseen denominaciones que incluyen metáforas en las unidades subrayadas. A continuación, vamos a centrarnos en el término 彩绘鼠尾草 [cǎi huì shǔ wěi cǎo], que es la *Salvia viridis* en chino, la cual dispone de dos metáforas simultáneamente. Igual que la forma-

Denominación latina	UT china	Transcripción en pinyin	Traducción literal	Observación
<i>Antirrhinum majus</i>	金鱼草	[jīn yú cǎo]	hierba de pez dorado	Las flores son numerosas y pequeñas como si fueran los peces dorados.
<i>Sandersonia aurantiaca</i>	宫灯百合	[gōng dēng bǎi hé]	lirio de farol de seda	Es de la subclase Liliidae y tiene las flores de color amarillo o naranja como si fueran las campanas pequeñas de seda.
<i>Trachelium caeruleum</i>	夕雾草	[xī wù cǎo]	hierba de niebla de atardecer	Tiene las flores pequeñas con inflorescencias umbeladas y de color violado.
<i>Eucomis comosa</i>	凤梨百合	[fèng lí bǎi hé]	lirio de piña	Es de la subclase Liliidae. Son plantas bulbosas perennes con rosetas basales de las hojas con tallos cubiertos de flores en forma de estrella, similares a los de la piña (A-Z Encyclopedia of garden plants).

TABLA 2. Ejemplos de las UT que explican la parte anterior de metáfora

ción de 玉米佛塔 [yù mǐ fó tǎ], de la flor *Banksia coccinea*, en los mercados existen varias salvias que tienen un nombre más general, «鼠尾草» [shǔ wěi cǎo], que significa ‘hierba de cola de ratón’. La variedad *Salvia viridis* es muy diferente de otras salvias que suelen ser verdes: su corola es siempre rosada o violácea, como si estuviera pintada, y la parte «彩绘» [cǎi huì] exactamente se refiere al concepto de ‘pintura’ o ‘algo pintado’.

4.4 Análisis de campos metafóricos de las UT

Desde la TCT, el estudio de campos metafóricos se puede remontar a los trabajos de Vidal y Cabré (2005, p. 189) en una investigación de las combinaciones léxicas metafóricas en el ámbito del genoma humano. Las autoras afirman que la estructuración en campos metafóricos y el establecimiento de vínculos analógicos entre muchos de los elementos de los dos dominios que componen una metáfora permiten agrupar combinaciones léxicas especializadas de carácter metafórico y explicar los procesos de transferencia que subyacen en cada una de ellas. Antes, autores como Kövecses (2002, p. 4) y Fernández Leborans (1977, p. 36) utilizan *dominio conceptual* y *campo semántico* para denominar el ámbito semántico o temático en el que se agrupan y se asocian las unidades terminológicas creadas por metáfora, a partir de las relaciones de analogía que se establecen entre unas y otras, que parten siempre de la existencia de un rasgo semántico compartido (Suárez Gallo, 2019, p. 209).

Además, en la teoría de la metáfora conceptual, Lakoff y Johnson defienden que el lenguaje y el pensamiento están ligados y se estructuran de acuerdo con la experiencia corpórea (*embodied experience*). Así, la experiencia perceptual y espacial son la base para categorizar el mundo, de manera que estas catego-

rias, que pertenecen a dominios concretos, se hacen corresponder con dominios más abstractos, como los de la emoción, el tiempo o la estructura. Por lo tanto, la metáfora queda definida como:

Un conjunto de correspondencias (*conceptual mapping*) entre un dominio fuente (más concreto) y un dominio meta (más abstracto) y la función de la metáfora es ayudar a pensar y razonar sobre el dominio meta en función de la estructura y lógica del dominio fuente. Así, la metáfora deja de considerarse, a diferencia de lo que se postula en la tradición clásica, como un elemento embellecedor de la lengua y se ve como un instrumento esencial en la conceptualización del mundo que nos rodea. (Ruiz de Mendoza Ibáñez y Masegosa, 2010, p. 107)

Se considera que dos términos son análogos cuando se refieren a entidades físicas o abstractas que en parte son iguales y en parte diferentes, ya sea en su ser (forma, estructura interna, relaciones con otros seres) o por la impresión que producen. Admitido este presupuesto (la igualdad parcial de dos referentes, objetiva o subjetiva), el conocimiento de uno de ellos, generalmente empírico, puede permitir el acercamiento y el conocimiento, al menos parcial, del otro, que puede ser también empírico o puede ser ficcional, mental, sentimental, intuitivo, onírico, etc. (Bobes Naves, 2004, p. 127).

Entre las UTM que conseguimos, los dominios fuente provienen de varios ámbitos y estos campos suelen tener relaciones más cercanas a nuestra vida cotidiana. Por lo tanto, agrupamos los dominios fuente entre varios campos metafóricos: zoología, utensilios, humanidad (leyenda, historia), cuerpo celeste, alimentación y cuerpo humano, que se ilustran en la tabla 3.

Campos metafóricos	Ejemplos
zoología	金鱼草 [kīn yú cǎo] ('hierba de pez dorado'); 鼠尾草 [shǔ wěi cǎo] ('hierba de cola de ratón'); <i>Delphinium</i> ; <i>Fritillaria meleagris</i> 'alba'; 袋鼠爪 [dài shǔ zhǎo] ('garra de canguro').
utensilios	flor de cera; espuela de caballero; 蕾丝 [lěi sī] ('encaje'); 剑兰 [jiàn lán] ('orquídea de espalda')
humanidad (leyenda, historia)	帝王 [dì wáng] ('reino'); 皇后 [wáng hòu] ('reina'); 公主 [gōng zhǔ] ('princesa'); 伯利恒之星 [bó lì héng zhī xīng] ('estrella de belén'); narciso 'estrella de Belén'
cuerpo celeste	满天星 [mǎn tiān xīng] ('cielo de estrella'); 金光菊 [jīn guāng jú] ('crisantemo de luz dorada'); 蓝星球 [lán xīng qiú] ('planeta azul'); <i>aster</i>
alimentación	凤梨百合 [fèng lí bǎi hé] ('lirio de piña'); flor de arroz; 橙菠萝 [chéng bō luó] ('piña naranja')
cuerpo humano	掌 [zhǎng] ('palma de la mano'); 鸿运当头 [hóng yùn dāng tóu] ('suerte toca la cabeza')

TABLA 3. Clasificación de las UTM según los campos metafóricos

4.5 Conceptualización específica: el traspaso de denominaciones de otras plantas

Además de los campos metafóricos mencionados en el apartado anterior, donde los conceptos son más familiares y cercanos en nuestra vida cotidiana y que corresponden más específicamente a la definición de la metáfora (‘transferencia, transporte’), tanto en chino como en la lengua española encontramos traslados desde denominaciones de plantas más comunes para conceptualizar plantas menos familiares. En esta relación, podemos observar dos grupos diferenciados:

- La planta fuente y la planta meta comparten relación taxonómica.
- La planta fuente y la planta meta comparten características parecidas (forma, olor, color, etc.).

Para concretar este fenómeno, podemos aprovechar el ejemplo de la flor *Convallaria majalis*. Además de su nombre más conocido, *muguete*, esta flor posee otro nombre, que es *lirio de los valles*. En realidad, cuando la observamos, no entendemos por qué tiene relación con el lirio, ya que esta flor tan pequeña con las flores blancas en forma de campanillas no se parece a los lirios tradicionales. Si, en cambio, buscamos su taxonomía en los libros de botánica, nos damos cuenta de que esta flor es una de las plantas procedentes de la subclase *Liliidae*, por lo cual no resulta raro que le dieran el nombre relacionado con el lirio.

En el idioma chino también podemos verificar este fenómeno con ejemplos tal como se muestra en la tabla 4.

Como puede verse, las cuatro flores de la tabla 4 provienen de la familia *Orchidaceae* y en lengua china comparten el mismo morfema «兰» [lán], que significa ‘orquídea’ y que compone un campo semántico, que ayuda a entender, con más facilidad, el resto de las denominaciones complejas.

Otro ejemplo interesante en lengua china es que se suelen utilizar las flores domésticas o más familiares para denominar las flores extranjeras, o las que tengan denominaciones más abstractas, a pesar de que no compartan ninguna relación taxonómica, como muestran los ejemplos de la tabla 5.

En la lengua española también existe este tipo de conceptualización léxica. Por ejemplo, entre las denominaciones vulgares de *girasol*, se encuentra *maíz de teja*, aunque el *girasol* y el *maíz* no pertenecen a la misma familia.

En el idioma chino este tipo de lexicalización es especialmente frecuente entre los nombres vulgares y dialectos. Si reflexionamos sobre la causa, este fenómeno puede tener la misma función que la metaforización: ayudar a pensar y razonar sobre el dominio meta en función de la estructura y la lógica del dominio fuente.

Denominación latina	Denominación china en el mercado floral	Transcripción en pinyin	Traducción literal
<i>Phalaenopsis</i>	蝴蝶兰	[hú dié lán]	orquídea de mariposa
<i>Paphiopedilum</i>	兜兰	[dōu lán]	orquídea de bolsillo
<i>Cymbidium</i>	大花惠兰	[dà huā huì lán]	orquídea bella de flor grande
<i>Dendrobium bigibbum</i>	洋兰	[yáng lán]	orquídea occidental

TABLA 4. UT encontradas en los mercados chinos que comparten el mismo morfema núcleo «兰» [lán]

Denominación latina	Denominación china en el mercado floral	Transcripción en pinyin	Traducción literal	Relación taxonómica
<i>Ranunculus asiaticus</i>	洋牡丹	[yáng mǔ dān]	peonía occidental	<i>Ranunculus</i> (<i>Ranunculaceae</i>) Peonía (<i>Paoniaceae</i>)
<i>Dicentra spectabilis</i>	荷包牡丹	[hé bāo mǔ dān]	peonía de bolsa pequeña	<i>Dicentra</i> (<i>Papaveraceae</i>) Peonía (<i>Paoniaceae</i>)
<i>Chamelaucium uncinatum</i>	澳洲腊梅	[ào zhōu là méi]	<i>Chimonanthus</i> de cera australiano	<i>Chamelaucium</i> (<i>Myrtaceae</i>) <i>Chimonanthus</i> (<i>Calycanthaceae</i>)

TABLA 5. UT que contienen traspasos de denominaciones de otras plantas sin relación taxonómica

4.6 La transparencia metafórica

Una metáfora ofrece siempre la visión parcial de un objeto. No todas las metáforas, por más que iluminen alguna propiedad del concepto metaforizado, son igualmente claras. Es decir, existen metáforas que revelan más, o que esconden menos que otras, de modo que la transparencia y la familiaridad de las metáforas son diferentes. Aquí, utilizamos el término *transparencia metafórica* a fin de reflexionar sobre la relación entre las unidades metafóricas y los usuarios.

Como se ha presentado anteriormente, los dominios fuente de las unidades metafóricas pueden ser diversos, y algunos de ellos provienen de nuestra vida cotidiana o incluso se refieren a las partes de nuestro cuerpo. Sin embargo, los dominios que tienen origen en personajes históricos o leyendas o que han sido transferidos desde otros idiomas pueden ser menos familiares.

Por ejemplo, los casos *flor de arroz*, *flor de cera*, *conejito*, 凤梨百合 [fèng lí bǎi hé] ('lirio de piña'), 蝴蝶兰 [hú dié lán] ('orquídea de mariposa') o 剑兰 [jiàn lán] ('orquídea de espalda') poseen dominios fuente que son más cercanos a nuestra vida cotidiana y los dominios fuente y los dominios meta comparten factores de semejanza fácilmente detectables. No obstante, en casos como 伯利恒之星 [bó lì héng zhī xīng] ('estrella de Belén'), para saber qué relación hay entre la flor y la estrella de Belén debemos conocer la leyenda de Jesucristo. En el caso de 荷包牡丹 [hé bāo mǔ dān] ('peonía de bolsa pequeña'), para conocer la relación metafórica tenemos que saber qué es esta bolsa pequeña tradicional de China. Igualmente, para el término *espuela de caballero* hace falta que pensemos cómo son las espuelas de los caballeros y por qué esta flor se parece a la espuela específica de los caballeros y no a una espuela ordinaria. Aunque *estrella de Belén*, 荷包 [hé bāo] y *espuela de caballero* provienen de dominios fuente que son concretos, no son conceptos comunes y transparentes (como la mariposa, la piña, el conejito) y precisan más tiempo de procesamiento.

Por añadidura, para los locales, las unidades metafóricas provenientes de otros idiomas suelen ser las metáforas menos transparentes. Como se ha mencionado en el apartado anterior para la palabra *Delphinium*, que se refiere al delfín, y para *Fritillaria meleagris*

'alba', en que *meleagris* significa 'pavo', en los mercados chinos también existen otros términos que tienen origen en lenguas extranjeras, como se ilustra en la tabla 6.

Podemos ver que hay préstamos directos del japonés, ya que los caracteres en ambos idiomas son casi los mismos, salvo la diferencia entre caracteres chinos tradicionales y simplificados.⁹ La denominación de la flor *Centaurea cyanus* en chino es 矢车菊 [shǐ chē jú], pero la palabra «矢车» [shǐ chē] no proviene del chino. «矢車» en japonés significa 'molino de viento', ya que las flores suelen ser de color azul intenso y se disponen en capítulos de 1,5-3 cm de diámetro, con un anillo de unas pocas flores liguladas grandes y extendidas que rodean un grupo central de flores discoidales que son muy parecidas a los molinos de viento. Además, el término para la flor *Spiraea cantoniensis* también proviene del japonés: «手球» en chino corresponde a dos sentidos, el balonmano y la falta en un partido de fútbol. No obstante, en japonés, entre los sentidos de «小手毬», se encuentra uno que es 'bolitas tejidas de hilo', lo cual coincide con las flores blancas agrupadas de la *Spiraea*.

La unidad 玉米百合 [yù mǐ bǎi hé] de la flor *Ixia*, en realidad, tiene su origen en su denominación en inglés, *corn lily*. Esta flor presenta tres estambres, el ovario es ínfero, el estilo se halla dividido en tres ramas filiformes. Las flores se disponen en racimos o espigas terminales y su forma agrupada se parece mucho a la del maíz. El arte del arreglo floral llegó a Japón de la mano de los monjes budistas, que lo aprendieron en China (Belcher, 1993). Aunque China tiene una historia muy larga en el arte de arreglos florales, en la actualidad, los países occidentales, especialmente Holanda y Alemania, son más potentes tanto en el desarrollo del cultivo de flores cortadas como en la teoría y la formación de floristas. Cuando los productos florales entran en el mercado chino, la traducción desempeña un papel sumamente importante, ya que el uso del latín o del inglés no es funcional en China, de un modo general. Por lo tanto, las flores no domésticas suelen tener nombres traducidos. Si queremos estandarizar UT de flores cortadas en el mercado internacional, como las situaciones en Europa y China son distintas, deberemos adoptar estrategias adecuadas a cada situación.

Denominación latina	Denominación encontrada en los mercados chinos	Transcripción en pinyin	Traducción literal	Término original
<i>Centaurea cyanus</i>	矢车菊	[shǐ chē jú]	crisantemo de molino de viento	ヤグルマギク (矢車菊)
<i>Spiraea cantoniensis</i>	小手球	[xiǎo shǒu qiú]	bolita de mano	コデマリ (小手毬)
<i>Ixia</i>	玉米百合	[yù mǐ bǎi hé]	lirio de maíz	corn lily

TABLA 6. UT metafóricas procedentes de lenguas extranjeras en mercados florales chinos

5 Conclusiones

El sector emergente de las flores cortadas, con un desarrollo bastante rápido, cuenta con productos florales y conceptos nuevos, y por lo tanto con neologismos abundantes. Algunos escenarios comunicativos (por ejemplo, las subastas de flores cortadas) requieren utilizar terminología precisa para identificar los diferentes productos y transmitir conocimiento especializado, como la que puede encontrarse en los libros de arreglos y diseños florales. Entre estas UT destacan las unidades de carácter metafórico.

Desde la TCT (Cabr , 1999), se considera que la presencia y la funcionalidad de las met foras en la terminolog a es fundamental porque forman parte del lenguaje natural. La met fora no solamente es una cuesti n formal o denominativa, sino que corresponde a los procesos cognitivos, ya que las UT son unidades de naturaleza interdisciplinaria (cognitivas, ling isticas y sociales). En este estudio, se han presentado unidades metaf ricas concretas del  mbito de las flores cortadas en chino y en espa ol, usadas en entornos profesionales, cient ficos y de comercio internacional.

Con los ejemplos seleccionados, hemos podido comprobar que la met fora puede estar ubicada en unidades terminol gicas monol xicas o polil xicas, y que entre estas hay met foras parciales ubicadas en posiciones distintas dentro de los sintagmas o unidades polil xicas. Incluso, un t rmino puede contener dos met foras al mismo tiempo.

Los campos metaf ricos de alimentaci n, zoolog a, cuerpo humano, cuerpos celestes e, incluso, leyendas e historias revelan que los seres humanos utilizamos met foras para acercarnos a conceptos abstractos y complicados a trav s de conceptos m s familiares. Podemos encontrar la semejanza entre el uso de estos campos metaf ricos tanto entre los usuarios chinos como entre los espa oles, aunque, por a adidura, estas met foras pueden mostrar diferencias en el nivel de transparencia y de conceptualizaci n. Otro tipo de transferencia semejante a la met fora consiste en el uso de denominaciones de flores populares y conocidas para denominar flores provenientes de otros pa ses o con denominaciones m s complejas o menos transparentes.

Adem s, el an lisis sobre el uso real de las UT en los mercados espa oles y chinos nos aporta varias claves rese ables para este estudio. Por ejemplo, a causa del uso amplio del lat n en las unidades utilizadas en los mercados espa oles, los t rminos son m s unificados y estables. Asimismo, en Espa a existen m s variantes interling isticas debido a que se pueden encontrar denominaciones inglesas, latinas, espa olas y catalanas. Por otra parte, en el caso de China, observamos que el uso de la met fora en las denominaciones usuales en los mercados de flores cortadas es mucho m s frecuente, lo que nos abre l neas de trabajo futuras sobre su funcionalidad. ✿

6 Bibliograf a

- BELCHER, Betty (1993). *Creative flower arranging: Floral design for home and flower show*. Portland, OR: Timber Press.
- BLACK, Max (1962). *Models and metaphors: Studies in language and philosophy*. Ithaca, Nueva York: Cornell University Press.
- BLACKWELL, Alan F. (2006). «The reification of metaphor as a design tool». *ACM Transactions on Computer-Human Interaction (TOCHI)* [Nueva York: Association for Computing Machinery], 13 (4), pp. 490-530.
- BLOOM'S (2017). *BASICS plant flashcards: Cut flowers*. Ratingen: Bloom's GmbH.
- BLOOM'S. (2018). «100 xianqiehua (100 鲜切花)». [100 cut flowers], vol.1 & 2. (Wen, Trad.) Zhongyuan Nongmin Chubanshe (中原农民出版社) [Editorial de Zhongyuan Nongmin].
- BOBES NAVES, Mar a del Carmen (2004). «La expresi n metaf rica: su sentido». En: *La met fora*. Madrid: Gredos, pp. 127-128.
- CABR , Maria Teresa (1993). *La terminolog a: Teor a, metodolog a, aplicaciones*. Barcelona: Ant rtida: Emp ries.
- (1999). *La terminolog a: Representaci n y comunicaci n*. Barcelona: Universitat Pompeu Fabra. Institut Universitari de Ling  stica Aplicada.
- (2002). «Terminolog a y ling  stica: la teor a de las puertas». *Estudios de Ling  stica del Espa ol*, 16, p. 3.
- (2009). «La teor a comunicativa de la terminolog a, una aproximaci n ling  stica a los t rminos». *Revue Fran aise de Linguistique Appliqu e*, XIV, p. 9-15.
- (2010). *Terminolog a y normalizaci n ling  stica*. Barcelona: Universitat Pompeu Fabra.

- CABRÉ, Maria Teresa; ESTOPÀ, Rosa (2005). «Unidades de conocimiento especializado, caracterización y tipología». En: CABRÉ, Maria Teresa; BACH, Carmen (ed.). *Connexement, llenguatge i discurs especialitzat*. Barcelona: Documenta Universitaria, pp. 69-94.
- CAO, Yumei (曹玉美) (2012). *Illustrated color landscape plants*. Beijing: China Forestry Publishing House.
- CHEN HUBIAO et.al (2012). *Medicinal Plant Images Database*. Hongkong: Hong Kong Baptist University, School of Chinese Medicine. [en línea] <<https://library.hkbu.edu.hk/electronic/libdbs/mpd/>> [Consulta: marzo, 2023]
- CIAPUSCIO, Guiomar Elena (2011). «De metáforas durmientes, endurecidas y nómadas: un enfoque lingüístico de las metáforas en la comunicación de la ciencia». *Arbor*, 187 (747), pp. 89-97.
- COROMINES Joan; PASCUAL, José Antonio (1980). *Diccionario crítico etimológico castellano e hispánico*. Madrid: Gredos.
- «Delphinium». En: *Flora vascular*. [en línea]. <<https://www.floravascular.com/index.php?genero=Delphinium>> [Consulta: 18 agosto 2022].
- FERNÁNDEZ LEBORANS, María Jesús (1977). *Campo semántico y connotación*. Madrid: CUPSA.
- GONG XUEMEI (龚雪梅) et al. (2012). «Huahui Yuanyi (花卉园艺)» [Flores y jardinería] Beijing: Jixie Gongye Chubanshe (机械工业出版社) [China Machine Press]
- GÜNTHER HAENSCH et al. (1993). *Nuevo diccionario de americanismos, Tomo I-Nuevos diccionarios de colombianismos*. Santafé de Bogotá: Instituto Caro y Cuervo.
- GRADY, Joseph (1997). *Foundations of meaning: Primary metaphors and primary scenes*. Tesis doctoral. Berkeley: University of California.
- JIANG, Shengde (江胜德) (2013). *1000 of garden plants illustrations*. Beijing: China Forestry Publishing House.
- KÖVECSES, Zoltán (2002). «Emotion concepts: Social constructionism and cognitive linguistics». En: FUSSELL, Susan R. (ed.). *The verbal communication of emotions*. Mahwah, NJ: Lawrence Erlbaum Associates, pp. 117-132.
- LAKOFF, George; JOHNSON, Mark (1995). *Metáforas de la vida cotidiana*. Madrid: Cátedra.
- (2008). *Metaphors we live by*. Chicago: University of Chicago Press.
- LI, Baoli; HU, Bingyi (李葆莉, 胡炳义) (2015). «Guojia Yaodian Zongyao Caise Tujian (国家药典中药彩色图鉴)» [Diccionario nacional de medicina tradicional china] Beijing: Junshi Yixue Kexue Chubanshe (军事医学科学出版社) [Military Medical Science Press].
- LI, Qin; CHEN, Defu (李琴, 陈德富) (2019). «Woguo Xianqiehua Chukou Fazhan de Xianzhuang Wenti yu Duice (我国鲜切花出口发展的现状、问题与对策)» [The current situation, problems and countermeasures of China's cut flower export development]. Vol. 7: *Duiwai Maoyi Shiwu (对外贸易实务)* [Practice in Foreign Economic Relations and Trade], pp. 46-49.
- LI YINPU (李印普) (2017). «Huahui Baodian (花卉宝典)» [Diccionario de flores] Beijing: China Forestry Publishing House.
- OLIVEIRA, Isabelle (2005). «La metáfora: estrategia de aprendizaje y sistema de conocimiento». En: *Comunicar y enseñar a comunicar el conocimiento especializado: Actas de la V jornada-coloquio de AETER*. Madrid: Instituto Cervantes, pp. 149-162.
- (2009). *Nature et fonctions de la métaphore en science: L'exemple de la cardiologie*. París: Editions L'Harmattan.
- PACKARD, Jerome (2000). *The morphology of Chinese: A linguistic and cognitive approach*. Cambridge: Cambridge University Press, pp. 14-25.
- PRELLI, Lawrence (1989). *A rhetoric of science: Inventing scientific discourse*. Columbia: University of South Carolina Press.
- Real Academia Española (2014). *Diccionario de la lengua española* (23.^a ed.). [en línea] <<https://dle.rae.es>> [Consulta: marzo de 2023].
- ROSSI, Micaela (2017). «Terminological metaphors and the nomadism of specialised terms». En: *Multiple perspectives on terminological variation*. Ámsterdam; Filadelfia: John Benjamins, pp. 181-212.
- RUIZ DE MENDOZA IBÁÑEZ, Francisco José; MASEGOSA, Alicia Galera (2010). «Mecanismos cognitivos en la conceptualización del mundo. La metáfora». *Cuadernos de Neuropsicología / Panamerican Journal of Neuropsychology*, 4 (2).
- SUÁREZ GALLO, Miriam (2011). «La metáfora en la construcción del discurso especializado y en la divulgación científica». En: *Análisis en el discurso del genoma humano*. Proyecto de tesis doctoral. Barcelona: Universitat Pompeu Fabra. Institut Universitari de Lingüística Aplicada.
- (2019). *La metáfora en el discurso especializado del genoma humano*. Tesis doctoral. Barcelona: Universitat Pompeu Fabra.
- TEMMERMAN, Rita (2021). «Metaphorical models and the translator's approach to scientific texts». *Linguistica Antverpiensia: New Series: Themes in Translation Studies*, [Amberes], 1 (octubre), pp. 211-226.

- VIDAL, Vanesa; CABRÉ, Maria Teresa (2005). «La combinatoria léxica en la enseñanza y aprendizaje de lenguas para propósitos específicos». En: CASTILLO CARBALLO, M. A.; CRUZ MOYA, O., GARCÍA PLATERO, J. M.; MORA GUTIÉRREZ, J. P. (coord.). *Las gramáticas y los diccionarios en la enseñanza del español como segunda lengua: deseo y realidad* (Actas del XV Congreso Internacional de ASELE: Sevilla). Sevilla: Universidad de Sevilla, pp. 896-902.
- WANG, Yicheng (王意成) (2018). «1000 Zhong Huahui Jianshang Tujian (1000种花卉鉴赏图鉴)» [Diccionario de 1000 flores] Beijing: Zhongguo Shuili Shuidian Chubanshe (中国水利水电出版社) [China Water & Power Press].
- WU, Zheng Yi et.al (1983). *Flora reipublicae popularis sinicae*. Beijing: Science Press.
- XINHUA CISHUSHE (新华辞书社). 2022. «Xinhua Zidian (新华字典)». [Diccionario de Xinhua] (12.ª ed.) Beijing: Shangwu Yinshuguan (商务印书馆) [The Commercial Press]

Notas:

1. Las flores cortadas son flores o capullos de flores (a menudo, con algunos tallos y hojas) que se han cortado de la planta que las produce (Li y Chen, 2019, p. 46). Suelen retirarse de la planta para uso decorativo. Los usos típicos son exhibiciones de jarrones, coronas, guirnaldas y ramos de novia.
2. Por ejemplo, en el genoma humano, las unidades terminológicas metafóricas (UTM) encontradas provienen de campos tales como ciencias del lenguaje (traducción, transcripción), militar (diana), información (código genético), etc.
3. Bloom's es una empresa alemana de fama mundial cuyos libros actualmente sirven con mucha frecuencia como recursos educativos importantes para los floristas de todo el mundo.
4. Terminology Online (termonline.cn), creada y organizada por el Comité Nacional de Terminología en Ciencia y Tecnología de la República Popular China, se posiciona como una plataforma donde se ofrecen servicios públicos sobre el conocimiento terminológico.
5. Por ejemplo, para la gran cantidad de rosas existentes, el BASICS plant flashcards solo nos brinda dos tipos generales: *rosa cultivar (solitary flowerhead)* y *rosa cultivar (multiple flowerheads)*. Para un estudio más detallado, en este artículo también analizamos las unidades encontradas en los mercados que no se incluyen en BASICS plant flashcards, como *rosa ramificada 'cappuccino' flor de arroz*, etc.
6. WANG (2002 [1953]), citado por Shuxiang Lü, por ejemplo, además de utilizar criterios semánticos (que generalmente corresponden a la «unidad básica de significado»), define la *cí* en chino como «la parte más pequeña de la lengua que se puede utilizar de forma independiente» o «la parte de la frase que se puede utilizar de forma independiente».
7. Consorte Yu (muerta en 202 aC), también conocida como Yu la Bella, fue una concubina del caudillo Xiang Yu, quien compitió con Liu Bang (emperador Gao), fundador de la dinastía Han por la supremacía en China durante la contienda Chu-Han (206-202 aC).
8. En realidad, la flor 腊梅 [là méi] corresponde a *Chimonanthus praecox*; en las traducciones realizadas sobre cultura china, esta flor famosa también se traduce como *ciruelo chino*. No obstante, *ciruelo* pertenece a la familia *Rosaceae*.
9. Los caracteres chinos simplificados (简体中文 [jiǎn tǐ zhōng wén], según la norma ISO 15924) son el conjunto de caracteres de escritura del idioma chino resultantes del proceso de simplificación de los caracteres chinos llevado a cabo por el Gobierno de la República Popular China durante la segunda mitad del siglo XX.

TERMINÀLIA parla amb...

Ramon López de Mántaras i Badia, director de l'Institut de Recerca en Intel·ligència Artificial



Ramon López de Mántaras i Badia Sant Vicenç de Castellet, 1952

Ramon López de Mántaras és doctor en Física per la Universitat Paul Sabatier de Tolosa i doctor en Informàtica per la Universitat Politècnica de Catalunya. Va fer estudis a la Universitat de Mondragón i a la Universitat de Califòrnia. És professor del Consell Superior d'Investigacions Científiques (CSIC), on dirigeix l'Institut de Recerca en Intel·ligència Artificial, ubicat al campus de la Universitat Autònoma de Barcelona.

És un dels pioners de la intel·ligència artificial a Catalunya i a Espanya. Ha rebut diversos premis: el Premi Ciutat de Barcelona d'Investigació (1982), el Premi Robert S. Engelmores de l'Associació Americana d'Intel·ligència Artificial (2011), el Premio Nacional de Informática de la Sociedad Científica Informática d'Espanya (2012), l'EurAI Distinguished Service Award de l'Associació Europea d'Intel·ligència Artificial (2016), el

Donald E. Walker Distinguished Service Award de la International Joint Conferences on Artificial Intelligence (2017) i el Premio Nacional de Investigación Julio Rey Pastor en Matemáticas y Tecnologías de la Información y las Comunicaciones (2018). L'any 2016 va ingressar com a membre numerari a l'Institut d'Estudis Catalans.

Els productes d'intel·ligència artificial (IA) com el ChatGPT s'han fet molt populars darrerament. La possibilitat d'interactuar-hi amb llenguatge natural, com ho faries xerrant amb una altra persona, probablement ha contribuït a la seva ràpida difusió. De fet, el ChatGPT es defineix com un *large language model* (LLM). També n'hi ha que el defineixen com un *large multimodal model* (LMM). Què és un *language model* (LM) i un *large language model*, i quina és la diferència entre aquests models i els *large multimodal models*?

Un LM és un model computacional de llenguatge que es basa en el càlcul de la probabilitat de seqüències de paraules. És a dir, calcula la probabilitat que unes paraules vagin a continuació d'unes altres en un idioma determinat. Un LM s'entrena només a partir de text. En canvi, un LLM és un model computacio-

nal (basat en xarxes neuronals artificials profundes) que s'ha entrenat amb un corpus enorme i multimodal (text, imatges, veu), i això fa que l'avaluació probabilística de quines paraules segueixen unes altres paraules sigui més acurada. A més de text, els LLM poden generar imatges basant-se en descripcions textuales o imatges introduïdes per l'usuari, i també poden generar textos molt més llargs que els LM i en diversos idiomes.

Quines aplicacions tenen els LLM en la vida quotidiana?

Poden tenir diverses aplicacions. Les més importants són: bots per xatejar, traducció automàtica, generació de textos (notícies, informes financers, redacció de contractes, respostes a correus electrònics...), transcripció de veu a text, assistents virtuals, resum d'articles i documents, recomanació de productes

(música, pel·lícules). Tot i que s'estan aplicant en sectors com ara la medicina, cal anar molt amb compte quan es tracta d'aplicacions en què la precisió de la informació és crítica. Per exemple, en un estudi titulat «Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine», publicat al *New England Journal of Medicine* el 30 de març de 2023, els metges van avaluar la fiabilitat del GPT-4 formulant la mateixa pregunta diverses vegades durant diversos dies. Utilitzant mètriques comunes per avaluar la similitud de les respostes generades, es va trobar una similitud molt baixa i una gran variabilitat entre les respostes a la mateixa pregunta. Globalment, només el 41 % de les respostes de GPT-4 coincidien amb respostes reconegudes com a correctes en medicina.

Els LLM han tingut molta rellevància en aplicacions lingüístiques com la traducció automàtica neuronal (TAN)



Font: arxíu personal

així començar a aprendre a relacionar el significant amb el significat. És un problema obert des que existeix la IA (a mitjans dels anys cinquanta) i es coneix com el «problema de la base simbòlica» (*symbol grounding problema*, en anglès).

Com a lingüistes, sovint ens costa assimilar la idea que un LM i un LLM calculen o processen llengua. Què té d'especial la IA que proporciona o analitza textos en relació amb la IA que proporciona respostes en altres formats (imatges, dades numèriques, etc.)?

La IA és molt més que els LLM i l'aprenentatge profund (DL, *deep learning*). Hi ha molts algoritmes que tenen poc o res en comú amb els dels LLM. Per exemple, en visió per computador s'usen xarxes convolucionals per identificar objectes en imatges que són diferents dels que s'usen en els LLM. En jocs de tauler com ara el go o els escacs s'usen combinacions d'algoritmes basats en xarxes neuronals profundes i algoritmes d'aprenentatge per reforç que són també molt diferents. En planificació i raonament s'usen tècniques

Les tecnologies com la traducció automàtica o la generació creativa de text son molt lluny de tocar sostre

o la generació creativa de text. Com a tecnologia lingüística, han tocat sostre? Quina aplicació, present o futura, tenen?

No! Som molt lluny de tocar sostre. De fet, tot i l'aparent capacitat de comprensió del llenguatge, de fet no comprenen res, tret que acceptem una definició molt restrictiva i superficial del que vol dir comprendre. Si *comprendre* es limita a generar

cadenes de text que els humans trobem coherents, aleshores podríem dir que comprenen el llenguatge. Però, en la meua opinió, i d'acord amb el que deia Saussure, el llenguatge està format pel significant i el significat, i els LLM només tracten (s'entrenen) amb el significant i, per tant, no hi ha cap possibilitat que aprenguin el component del significat, ni semànticament ni pragmàticament ni culturalment parlant. Estic força d'acord amb l'experta en lingüística computacional Emily Bender quan diu que els LLM són «lloros estocàstics». En un futur, l'ús d'aquests models en IA corpòries (és a dir, en robots físics) pot ser una via per començar a superar la manca de coneixements de sentit comú (coneixements sobre el món), ja que una IA amb cos permetria experimentar amb el món i

Els LLM només s'entrenen amb els significants i no hi ha cap possibilitat que aprenguin el component del significat, ni semànticament ni pragmàticament ni culturalment parlant

d'IA, basades en la lògica matemàtica, que també són completament diferents. En sistemes multiagent, en què els problemes es resolen mitjançant la col·laboració entre diversos agents intel·ligents, les tècniques d'IA també són molt diferents. En definitiva, en IA tenim un

ventall molt ampli d'algoritmes molt diversos que s'usen, individualment o en combinació, en funció del problema que cal resoldre. Per exemple, el famós sistema AlphaFold, que prediu el plegament de proteïnes, combina trenta-dos algoritmes diferents. No existeix un únic algoritme general (és a dir, una IA general) capaç de dur a terme qualsevol tasca.

Un altre aspecte que ens costa assimilar com a lingüistes és que pràcticament totes les dades són aptes per entrenar una IA. Per exemple, en traducció automàtica estadística (TAE) i en altres sistemes entrenats amb quantitats de dades inferiors a les utilitzades en un LLM, es tenia

Els LLM no tenen una comprensió real de la

llengua i del món com tenim nosaltres, els humans.

Només detecten patrons lingüístics comuns en el corpus d'entrenament

molt en compte la màxima *garbage in, garbage out*, que ve a dir que un sistema entrenat amb dades de poc valor proporcionarà resultats que també seran d'una qualitat baixa. Certament hem superat aquesta limitació?

No, no hem superat aquesta limitació. De fet és molt evident que els LLM continuen generant *garbage* i això és un problema molt seriós, ja que generen respostes esbiaixades pel que fa a aspectes com ara gènere, racisme... Aquestes respostes s'intenten evitar introduint filtres de manera manual, però això no és cap solució real, ja que només són pedaços que no podran solucionar mai un problema que és inherent a l'absoluta manca de comprensió del

llenguatge per part dels LLM. Això, juntament amb el problema dels múltiples errors factuais que generen, que fa que siguin uns grans productors de desinformació, pot causar polarització en la societat i posar en perill fins i tot la democràcia. Com molt bé diu Daniel Dennett, els sistemes com el ChatGPT generen «counterfeit people» i, si falsificar moneda és un crim, falsificar persones hauria de ser-ho també.

I una capacitat més, que sembla contradir alguns dels principis lingüístics (i particularment traductològics) anteriorment inamovibles: la generació de text en una llengua a partir de sistemes entrenats amb textos en altres llengües. Què fa que un sistema de generació de text com un LLM pugui produir text en una llengua que està molt poc representada en el seu corpus d'entrenament?

Els LLM han estat entrenats amb textos en moltes llengües diferents i dissenyats per detectar-hi patrons lingüístics, és a dir, relacions entre paraules, frases i estructures gramaticals. De fet, detectar patrons és el que realment saben fer les xarxes neuronals profundes. Com que aquests patrons són comuns a moltes llengües, els LLM els utilitzen per generar textos en llengües fins i tot poc representades en el conjunt d'entrenament. Ara bé, cal deixar clar que els resultats en llengües poc representades són de qualitat força inferior. Òbviament, en el cas de llengües que, a més d'estar poc representades en els corpus d'entrenament, tenen estructures lingüístiques molt diferents, els resultats seran molt deficients.

Parlem ara dels productes textuals d'aquests sistemes. N'hi ha que s'apropen amb reticències als productes, com el ChatGPT, a causa de l'homogeneïtat del text produït. Per exemple, qualsevol text analític o reflexiu que escrivia el ChatGPT3 acostumava a acabar amb un paràgraf introduït pel connector «Overall, ...». Aquests límits creatius són

insuperables? O arribarà el sistema que sigui capaç d'emular la manera de fer servir la llengua de cada usuari?

Tal com he respost en una de les preguntes anteriors, aquests models no tenen una comprensió real de la llengua i del món com tenim nosaltres, els humans. Només detecten patrons lingüístics comuns en el corpus d'entrenament. Per això, els textos que generen són homogenis. En teoria, es podria millorar el que fan actualment si s'entrenessin amb un corpus suficientment ampli específic per a cada usuari. Òbviament, llevat d'alguns autors, la immensa majoria dels usuaris no han generat un corpus prou ampli de textos

La recerca en terminologia, pot contribuir-hi

millorant, per exemple, la comprensió del llenguatge especialitzat

per entrenar un LLM. D'altra banda, aquesta personalització, amb els LLM actuals, seria molt costosa tant en temps com en termes econòmics. Caldrien models computacionals diferents dels LLM, que permetessin aprendre la manera de fer servir la llengua de cada usuari partint de corpus relativament petits.

Com a lingüistes, quines eines d'IA poden contribuir a la nostra recerca? I viceversa, quina recerca nostra pot contribuir al desenvolupament de la IA amb aplicacions de llengua o lingüística? En particular, la recerca en terminologia i els productes terminològics tenen cabuda en l'escenari futur de desenvolupament de la IA?

Crec que hi ha treballs que han desenvolupat molts lingüistes computacionals i també experts en *natural language understanding* (NLU) al llarg dels darrers cinquanta anys que



Foto: arxiu personal

podrien ser molt útils per millorar els models actuals. Per exemple, mitjançant sistemes híbrids que aprofitin els avantatges de les aproximacions antigues, basades en representacions computacionals del coneixement que atacaven els problemes de les dimensions semàntica, pragmàtica i cultural de la llengua, i les aproximacions actuals, basades en la detecció de patrons en dades reals. Aquesta hibridació potser permetria desenvolupar models de llenguatge més «personalitzables» (vegeu la pregunta anterior).

Pel que fa a la recerca en terminologia, pot contribuir-hi millorant, per exemple, la comprensió del llenguatge especialitzat aportant recur-

sos terminològics i diccionaris per entrenar i millorar els algoritmes.

I una darrera pregunta, relacionada amb la terminologia. La popularització del ChatGPT ha comportat també un vocabulari associat. Hi ha un terme en particular que s'utilitza molt: *prompt*. Ja es parla de *prompt engineering* i de *ser expert en prompts*. Què és un *prompt*? I com el traduiries al català?

És una entrada en forma de text (frase, paràgraf, pregunta...), més aviat breu, que s'utilitza per iniciar la generació d'un text per part del ChatGPT i també per orientar sobre quin tipus de text volem obtenir i de quin tema. Aquest text pot ser

una continuació de la frase o de les frases d'entrada o una resposta en cas que l'entrada sigui una pregunta. Com més òptim sigui el *prompt* millor serà la qualitat de la resposta generada, per això és important l'habilitat de l'usuari a l'hora de dissenyar el *prompt*, i d'aquí sorgeix això del *prompt engineering* per referir-se a l'habilitat de crear *prompts* efectius.

No tinc clar quina seria la millor traducció. N'he vist diverses opcions en la bibliografia científica, com *entrada*, *instrucció*, *consigna* o *indicació*.

CONSELL DE REDACCIÓ

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

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.



RAMON FERRER I CANCHO

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

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 ingents (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	$y(n) = a n^b \exp(-cn)$	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 «ferta» 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 gòlem

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 pampallugues de 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?

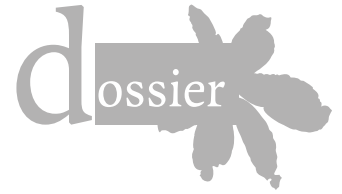
Entretant, comprendre el llenguatge continua sent el repte de la lingüística quantitativa, un objectiu compartit amb d'altres ciències i disciplines afins (Hernández-Fernández i Ferrer-i-Cancho, 2019), un repte convergent en què el ser i el saber estan per sobre de l'aparença, de la impostació tecnològica. 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 tecnologia*. 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; SEMPLÉ, 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 cuantitativa: 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 MANTARAS, 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-intel·ligencia-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/>.

El projecte AINA, la IA i les tecnologies del llenguatge

MARTA VILLEGAS MONTSERRAT

Barcelona Supercomputing Center

ORCID: 0000-0003-0711-0029

marta.villegas@bsc.es



Marta Villegas fa més de 25 anys que treballa com a investigadora en el camp del processament del llenguatge natural. Actualment és la responsable de la Unitat de Tecnologies de la Llengua al Barcelona Supercomputing Center - Centro Nacional de Supercomputación, on dirigeix els treballs per al desenvolupament de models lingüístics. La Unitat

ha compilat recentment el corpus espanyol i català més gran mai creat i ha desenvolupat models de referència *transformers* que han tingut un gran impacte, tant en el món acadèmic com en la indústria. Coordina el projecte AINA i és responsable de diversos projectes nacionals i europeus.

Resum

Una de les àrees més rellevants de la IA és el processament del llenguatge natural (PLN). En aquest àmbit, tot i que actualment la majoria dels grans models de llenguatge ja són multilingües, hi ha una diferència substancial entre les capacitats dels models pel que fa a l'anglès i a la resta de llengües. En aquest sentit, el projecte AINA té per objectiu desenvolupar la infraestructura necessària per què la inclusió del català a les aplicacions d'IA sigui prou atractiva i viable. Aquest article presenta els objectius del projecte i n'explica les característiques generals.

PARAULES CLAU: IA; tecnologies del llenguatge; català; PLN

Abstract

The AINA Project, Artificial Intelligence, and Language Technologies

One of the most relevant areas of AI is Natural Language Processing (NLP). In this area, even though most of the large language models are currently multilingual, there is an important difference between the capabilities of English models and the other languages. Thus, the AINA project aims at developing the necessary infrastructure so that the inclusion of Catalan in AI applications becomes appealing and feasible. This article presents the objectives of the project and explains its main characteristics.

KEYWORDS: AI; language technologies; Catalan; NLP

Una mica de context

En els darrers anys, la intel·ligència artificial (IA) ha estat causant sensació en el món de la tecnologia i el seu impacte en la nostra societat és cada vegada més i més gran. Sense cap mena de dubte, una de les àrees més rellevants de la IA és el processament del llenguatge natural (PLN), una disciplina que preten ensenyar les màquines a entendre i utilitzar el llenguatge humà.

Fins als anys 80, la majoria de sistemes de PLN es basaven en regles. A finals d'aquesta dècada, els algorismes d'aprenentatge automàtic es van fer cada vegada més populars i, aviat, els models estadístics van reemplaçar les regles, cosa que va suposar una millora substancial. La següent revolució va arribar el 2013 amb la introducció de les representacions vectorials de paraules (*embeddings*) com Word2vec,¹ GloVe² i FastText.³ Els *embeddings* encapsulen el «significat» d'una paraula en un vector després de llegir quantitats massives de text i analitzar els contextos en què apareix cada paraula. La idea és que paraules similars tenen contextos similars. Aquests primers models s'entrenaven amb una xarxa neuronal petita i tenien resultats prou bons, però no suficients, ja que cada paraula tenia un únic vector, quan en realitat una paraula pot tenir múltiples significats.

A finals del 2018, Google va publicar els models BERT introduint l'arquitectura *transformers*⁴ i provocant una autèntica revolució en el camp de les tecnologies del llenguatge. Des del primer moment, els *transformers* van suposar una millora espectacular en qualsevol de les tasques típiques del processament del llenguatge. Prova d'això són la quantitat i qualitat dels sistemes i aplicacions de PLN que van sorgir, com ara sistemes de traducció automàtica, assistents virtuals com Siri, Alexa o Google Assistant, sistemes de correcció i autocompletat de text, classificadors de documents, anàlisi de sentiments, entre molts d'altres. Google i les grans empreses tecnològiques nord-americanes van començar la cursa dels anomenats grans models de llenguatge (coneguts també com models fundacionals) en què la mida del model (mesurada en paràmetres), la quantitat de dades utilitzades per entrenar-lo i la quantitat de còmput necessari són cada vegada més

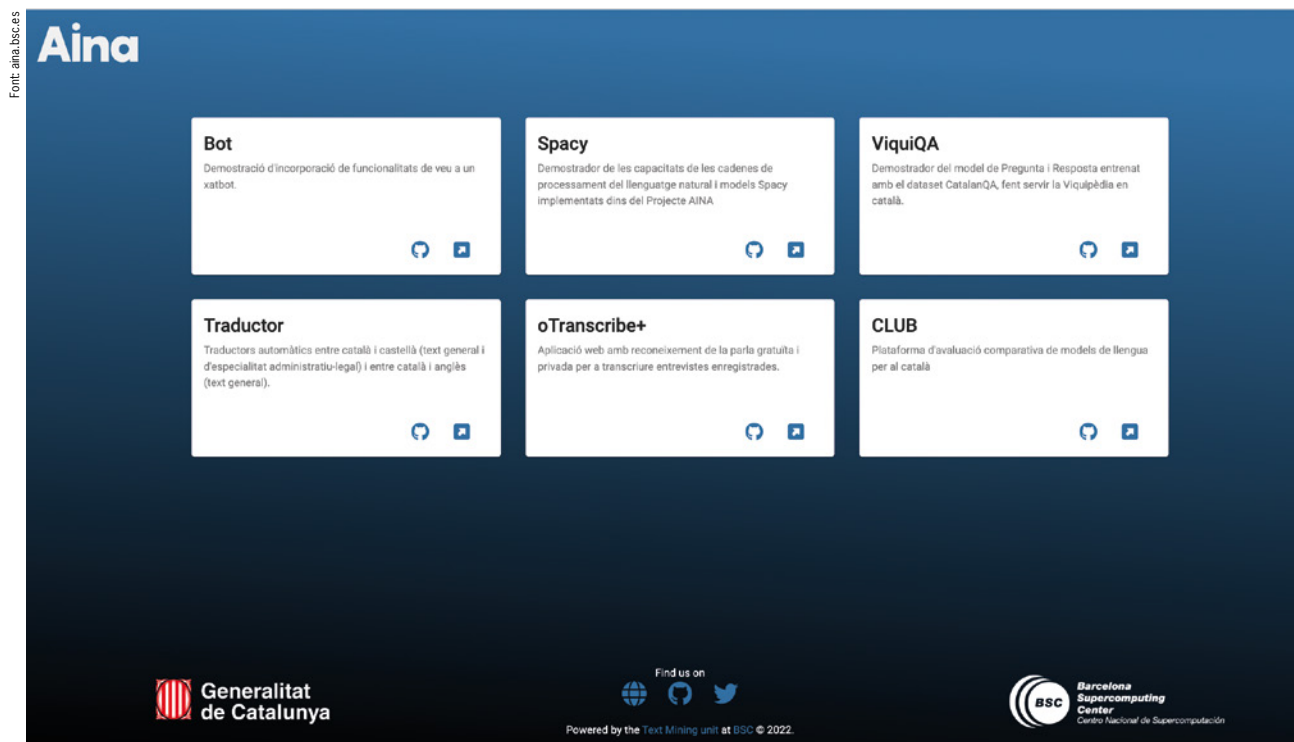
grans. Cal recordar que per entrenar el model GPT-3 d'OpenAI,⁵ que conté 175 mil milions de paràmetres, es van utilitzar 300 mil milions de *tokens* i es van necessitar 3.14e23 flops.⁶

Quan parlem de models del llenguatge, ens referim a models que aprenen la probabilitat d'ocurrència de les paraules, basant-se en exemples que han estat vistos. Disposar de bons models és crucial per a les tecnologies de la llengua (TL), ja que permet obtenir millors resultats i rendiments en qualsevol tasca de processament del llenguatge. Com ja hem vist, per a obtenir bons models cal disposar de grans quantitats de dades (*big data*) i capacitat de càlcul per poder processar-les.

Recentment, el xatbot ChatGPT, llançat per l'empresa nord-americana OpenAI al novembre del 2022, ha demostrat l'extrema capacitat disruptiva d'aquesta tecnologia, que canvia paradigmes i és capaç de millorar amb nous llançaments, com el model GPT4, també desenvolupat per OpenAI. La successió de models, arquitectures i propostes no para de créixer; així, al març del 2023, Google va publicar PaLM-E, «un model de llenguatge multimodal personificat».⁷ Poc després, Google va publicar Bard, la seva versió d'un xatbot amb IA. Gairebé al mateix temps, el gegant xinès dels motors de cerca Baidu va presentar el seu model Ernie.⁸

Sense cap mena de dubte, els xatbots impulsats per els grans models de llenguatge estan revolucionant la tecnologia de la informació tal com la coneixem, amb implicacions i impacte substancials per a la societat, la investigació i la indústria en general. ChatGPT i tecnologies similars desenvolupades per altres gegants tecnològics nord-americans o asiàtics es poden utilitzar com a assistents d'escriptura, ajudants personals, solucionadors de problemes en general, robots de text i, en general, *sparring* per a tasques, reptes i situacions quotidianes de la nostra vida personal o professional. Poden aplicar-se pràcticament a tots els àmbits, des de l'atenció al client fins a l'assistència sanitària, passant per la mobilitat, l'educació, les finances, les assegurances, el comerç electrònic i molts altres. Des del llançament de ChatGPT, milers de casos d'ús professional han estat ràpidament creats, implementats i validats per tercers, molts ja utilitzats en entorns de producció.





El cas és que la majoria d'aquests models, inicialment, es van generar per a l'anglès, és a dir, es van entrenar amb dades exclusivament, o molt majoritàriament, en anglès. Aviat, la necessitat de disposar de models per a altres llengües i la manca de dades suficients en aquestes van propiciar l'aparició dels primers models multilingües, models entrenats amb dades en diferents llengües que mostraven capacitats multilingües.

Tot i que podem afirmar que la majoria dels grans models de llenguatge són multilingües, el cert és que hi ha una diferència substancial entre les capacitats dels models pel que fa a l'anglès i a la resta de llengües. Els models són clarament anglocèntrics senzillament perquè s'han entrenat amb moltes dades en anglès i molt poques en altres llengües. L'escassa presència de la majoria de llengües en aplicacions de PLN provoca que moltes llengües mundials i europees puguin desaparèixer relativament aviat de l'ecosistema digital (primer pas per a la seva extinció en el món real).⁹ Això va alertar el Parlament Europeu, que el setembre de 2018 va aprovar una resolució per garantir l'equitat de les llengües en l'era digital.¹⁰

El projecte AINA

AINA és un projecte impulsat pel Departament d'Empresa i Treball que, amb la col·laboració del Barcelona Supercomputing Center, ha de permetre al català fer un salt qualitatiu i quantitatiu en l'ecosistema digital.

AINA neix en un moment especialment delicat, en què els avenços tecnològics en el camp del PLN són alhora una oportunitat i un repte per al català. D'una banda, les possibilitats que ofereixen les TL són cada vegada més extraordinàries i tenen un impacte més gran en les nostres vides. Però de l'altra, clarament es focalitzen en les llengües globals, deixant la resta en una situació de desavantatge.

AINA, doncs, té per objectiu desenvolupar la infraestructura necessària perquè la inclusió del català a les aplicacions d'IA sigui prou atractiva i viable, tant per a les grans companyies tecnològiques com per a la indústria local. Per tant, el projecte té com a objectiu el desenvolupament de recursos per a la IA/TL a fi d'aconseguir el següent:

- Proveir el català de la infraestructura necessària per al desenvolupament d'aplicacions basades en IA/TL (assistents de veu, traductors automàtics, agents conversacionals, etc.);
- Garantir que la inclusió del català a les aplicacions de IA/TL sigui rendible i atractiva per a les empreses del sector, tant a escala local com global;
- Impulsar el sector de la IA/TL a Catalunya millorant-ne la competitivitat;
- Aconseguir que el ciutadà de Catalunya pugui participar en català en el món digital al mateix nivell que els parlants d'una llengua global;
- Fomentar la investigació capdavantera en IA/TL per tal que Catalunya estigui ben preparada per respondre al repte digital i de la societat basada en el coneixement.

AINA és essencialment infraestructura lingüística en què el valor de les dades és cabdal: la tecnologia avança molt ràpidament, però les dades són persistents. Disposar de dades de qualitat suficients és un actiu segur i de futur que garanteix l'actualització de la tecnologia. D'altra banda, la IA/TL és sens dubte un àmbit en expansió que genera un immens volum de negoci i llocs de treball qualificats. Es calcula que al 2020 el mercat mundial de la IA/TL es va situar en 16.400 milions de dòlars i es preveu que arribi als 59.000 milions el 2027, amb un creixement anual del 20,1%. És, per tant, necessari garantir la competitivitat de Catalunya en aquest àmbit, mitjançant la vigilància tecnològica i sectorial i la recerca més puntera, amb la participació d'empreses i grups de recerca del sector, garantint l'enfortiment de tot l'ecosistema.

Quan parlem de dades lingüístiques ens referim a qualsevol dada de tipus textual o sonora (o audiovisual) produïda mitjançant el llenguatge humà. Les dades textuais serviran per entrenar models de llengua i desenvolupar aplicacions lingüístiques diverses, incloent-hi motors de traducció automàtica, i les dades de parla permetran crear sistemes de reconeixement i de síntesi de veu. Per tal de garantir el subministrament continu de dades, AINA preveu establir acords que defineixin i implementin protocols de publicació de dades, en compliment de les directives europees sobre dades obertes i reutilització de la informació del sector públic (RISP),¹¹ i respectant els principis FAI,¹² amb els grans proveïdors de dades. Cal fer les accions necessàries perquè els grans generadors de dades en català es constitueixin en subministradors de dades per a la IA/TL de manera sostinguda i programàtica. Per tal d'aconseguir-ho, s'estan abordant tots els aspectes necessaris incloent-hi aspectes tècnics i legals. Tècnicament, les accions han de permetre la publicació de dades que idealment permetin l'accés programàtic en formats processables. Per això és necessari el desenvolupament d'API o mecanismes de descàrrega/compartició de dades, definició de formats i metadades, etc. El primer exemple és l'accés a les dades de veu i text del Parlament de Catalunya que permet accedir a les dades de text i veu de les sessions del Parlament.

Part important són també les dades anotades; AINA ha invertit esforços considerables en la generació de dades anotades per a diferents tasques i s'està treballant en la generació d'un corpus d'instruccions que ens haurà de permetre «instruir» diferents models generatius.

Un altre aspecte fonamental en la infraestructura d'AINA són els models del llenguatge. Com ja hem vist, els models són una peça clau en el desenvolupament de noves aplicacions. La generació de models utilitzant tècniques d'aprenentatge profund i dades massives és una tasca extremadament costosa en recursos computacionals i humans, que difícilment

poden afrontar les empreses o els grups de recerca catalans. Últimament, la recerca i generació de models massius ha quedat en mans de les grans tecnològiques, que, tot i que cada vegada treballen més amb dades multilingües, solen deixar de banda les llengües no globals. Cal, doncs, garantir que l'ecosistema català disposi de models suficients. Així doncs, AINA vol generar models de la llengua en les arquitectures més novedoses i amb més impacte, incloent-hi models monolingües, multilingües i multimodals, prestant especial atenció a la participació en grans projectes multilingües (nacionals i internacionals) i garantint la presència del català en aquests models massius. En aquest sentit, actualment s'està treballant en la generació de tres models GPT3 trilingües (català, castellà i anglès) de diferents mides que després es podran adaptar a tasques i dominis específics.

A AINA, les tecnologies de la parla tenen especial rellevància. L'objectiu de les tecnologies de la parla és la implementació del mitjà més natural d'interacció com a interfície entre humans i màquines. La parla incorpora aspectes que no solem trobar en les formes escrites de la llengua com les variants dialectals i, fins i tot, emocions. La parla també aporta connotacions de trets personals com ara el gènere i l'edat. Per aquestes raons, la incorporació de la parla als productes tecnològics és molt important, tant des del punt de vista cultural com per fomentar l'ús de la llengua en les tecnologies, especialment per a llengües minoritzades com el català. Comparat amb les tecnologies de text, la incorporació de les tecnologies i funcionalitats de la parla té una barrera d'entrada, per la varietat de maneres de parlar, i també per la dificultat suplementària que introdueix el processament del senyal acústic. Per això, és imprescindible un esforç sostingut de captació de veus de tot l'àmbit geogràfic català. AINA ja va engegar la campanya de Common Voice al Principat, que s'ha estès a les Illes Balears i properament s'estendrà al País Valencià gràcies a la col·laboració del Projecte Vives, amb què el català ha passat a ser la segona llengua en nombre d'hores a la plataforma de referència. La generació de models de síntesi de la parla, el reconeixement de la parla, i la traducció automàtica parla a parla és una prioritat del projecte.

Les tecnologies de la traducció automàtica tenen també un paper clau a AINA. La traducció automàtica permet salvar les barreres idiomàtiques en entorns multilingües, tot garantint l'ús de la llengua pròpia. Això la converteix en una tecnologia clau per a la salvaguarda de les llengües minoritzades i de la diversitat lingüística. Entre altres funcionalitats essencials, destaquem el suport a la traducció documental, la traducció de continguts web, la subtitulació automàtica o la traducció en temps real. Per tal de dotar el català de les eines tecnològiques que permetin assolir aquestes funcionalitats AINA es proposa construir i posar a dis-



grans tecnològiques (p.ex. Google Translate), tot garantint la sobirania tecnològica i el control de l'accés a les dades, així com assegurar la presència del català dins de les plataformes més populars de programari lliure. Cal emfatitzar que tots els recursos generats per AINA es publiquen en obert i amb llicències el més permissives possible.

Finalment, un dels objectius d'AINA és posicionar Catalunya com a referent en l'àmbit de la IA/TL. El mercat de la IA/TL està en plena expansió. L'augment de la taxa d'adopció de la IA en els di-

versos sectors industrials i la demanda creixent de tecnologies com traducció i parla, així com l'augment de les àrees d'aplicació (p.ex. detecció de desinformació) impulsen el creixement del mercat. Catalunya, amb un ecosistema de recerca i teixit industrial propici, té el potencial per liderar la recerca i el desenvolupament en TL de llengües mitjanes i petites. Així doncs, AINA ha d'identificar els avenços de la tecnologia i respondre-hi mitjançant la vigilància tecnològica i identificar noves necessitats de les empreses i de la societat mitjançant la vigilància sectorial i de mercat. L'objectiu últim és propiciar l'adopció de les tecnologies innovadores, fomentar-ne l'adopció, la participació de les empreses del sector, la transferència tecnològica i l'exploració de nous àmbits d'aplicació. ✿

posició de la comunitat models de traducció basats en els últims avenços tecnològics entre el català i llengües d'interès, ja sigui per motiu de negoci (llengües europees, xinès, etc.) o proximitat geogràfica i social (aranès, llengües de la península, llengües de la immigració). Donada l'escassetat de dades existents per a la majoria d'aquests parells de llengües, està prevista la recerca i desenvolupament de tècniques d'entrenament no supervisat més adequades per a la traducció automàtica. Aquesta recerca és potencialment beneficiosa per a un gran nombre de llengües del món i contribuirà a posicionar la recerca catalana en aquest àmbit.

Com en altres casos, les tecnologies de traducció desenvolupades per AINA permetran disposar de solucions tecnològiques pròpies, independents de les

Notes:

1. Mikolov, Tomas; et al. (2013). «Efficient Estimation of Word Representations in Vector Space». arXiv:1301.3781
2. Pennington et al. (2014). «GloVe: Global Vectors for Word Representation».
3. Bojanowski et al. (2017). «Enriching Word Vectors with Subword Information»
4. Devlin J, Chang MW, Lee K, Toutanova K. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805. 2018 Oct 11
5. «Language Models are Few-Shot Learners» <https://arxiv.org/abs/2005.14165>
6. Un FLOP es una unitat de mesura de la potència de càlcul d'un ordinador que correspon a una operació de coma flotant per segon
7. <https://ai.googleblog.com/2023/03/palm-e-embodied-multimodal-language.html?m=1>
8. <https://www.aljazeera.com/economy/2023/3/16/chinas-baidu-unveils-chatgpt-rival-ernie>
9. [https://www.europarl.europa.eu/stoa/en/document/EPRS_STU\(2017\)598621](https://www.europarl.europa.eu/stoa/en/document/EPRS_STU(2017)598621)
10. <https://www.greens-efa.eu/en/article/press/victory-for-language-equality-in-the-european-parliament/>
11. https://administracionelectronica.gob.es/pae_Home/pae_Estrategias/pae_Gobierno_Abierto_Inicio/pae_Reutilizacion_de_la_informacion_en_el_sector_publico.html
12. <https://www.go-fair.org/fair-principles/>

Resiliència digital: aprofitar el poder del col·lectiu per preservar la llengua. Una història d'èxit per al català

ONA DE GIBERT BONET

Universitat d'Hèlsinki

ORCID: 0000-0002-7163-4807

ona.degibert@helsinki.fi



Ona de Gibert és estudiant de doctorat en Traducció Automàtica a la Universitat d'Hèlsinki, on contribueix al desenvolupament de tecnologies del llenguatge per a llengües amb pocs recursos. Graduada en Llengües i Literatures Modernes per la Universitat de Barcelona (2016), va obtenir un màster en Anàlisi i Processament del Llenguatge per la Universitat del País Basc - Euskal Herriko Unibertsitatea (UPV-EHU) (2018). El seu principal

camp de treball gira entorn de la traducció automàtica, des de la seva implementació pràctica a l'empresa fins a la recerca de nous mètodes en la seva posició actual. També es dedica a la divulgació i l'activisme científics per a la preservació de les llengües minoritzades en l'era digital. Ha treballat en el projecte AINA, en l'àmbit estatal, i en el projecte BigScience, en l'àmbit internacional.

Resum

El creixement generalitzat de la intel·ligència artificial (IA) ha fet que les tecnologies de la llengua siguin més accessibles que mai i ha portat aquesta tecnologia a la nostra vida diària. Tanmateix, el desenvolupament accelerat de la tecnologia lingüística comporta intrínsecament un biaix cap a una perspectiva anglocèntrica i eurocèntrica, que té com a resultat una representació i un reconeixement limitats de les llengües minoritzades. És ben sabut que la presència d'una llengua en línia en garanteix la supervivència. En aquest article explorem la història d'èxit del català, una llengua minoritzada amb una comunitat activa en línia, que ha establert les bases per al desenvolupament de les tecnologies de la llengua. La història d'èxit de la comunitat catalanoparlant demostra que les comunitats poden tenir un paper important en la supervivència i en l'evolució de les llengües minoritzades en l'era digital.

PARAULES CLAU: intel·ligència artificial; català; preservació de la llengua; codi obert

Abstract

Digital Resilience: Harnessing the Power of the Collective for Language Preservation. A Success Story for Catalan

The widespread growth of Artificial Intelligence (AI) has made language technology more accessible than ever before, bringing language technology into our daily lives. Nevertheless, the rapid development of language technology intrinsically carries a bias towards Anglo-centric and Euro-centric perspectives, leading to limited representation and recognition for minoritized languages. It is a fact that the online presence of a language ensures its survival. This article explores the success story of Catalan, a minoritized language with an active online community, which has laid the foundations for language technology development. The Catalan-speaking community's success story demonstrates how communities can play a significant role in the survival and evolution of minoritized languages in the digital age.

KEYWORDS: artificial intelligence; Catalan; language preservation; open-source

El ràpid avenç de la intel·ligència artificial (IA) en el segle XXI coincideix amb un desenvolupament imprevist de la tecnologia lingüística. El desenvolupament de grans models del llenguatge (LLM, *large language models*), sistemes neuronals de traducció automàtica (NMT, *neural machine translation*), aplicacions de reconeixement automàtic de la parla i d'assistents virtuals comercials estan obrint un nou espai de diàleg per al gran públic com no havia passat mai abans. Sembla que la tecnologia s'ha tornat més accessible a les masses i que és capaç de sortir-se'n de tasques difícils per a totes les llengües. Però, és realment així? El costat fosc del desenvolupament accelerat de la tecnologia lingüística comporta intrínsecament un biaix cap a una perspectiva anglocèntrica i eurocèntrica, cosa que dona com a resultat una representació i un reconeixement limitats de les llengües minoritàries.

Segons Bali et al. (2019), només entre deu i quinze llengües es veuen directament afavorides pels grans canvis en el paradigma de la IA; concretament, les llengües que tenen molts més recursos, com l'anglès, l'àrab i el castellà (vegeu la figura 1). A l'era digital en què vivim actualment, està demostrat que la presència d'una llengua a la Xarxa n'assegura la supervivència. En realitat, la distribució de llengües a Internet segueix una distribució de Zipf: hi ha molt poques llengües que hi apareixen molt (les llengües amb molts recursos) i, a l'inrevés, n'hi ha moltes que no hi són gaire presents. Més del 90 % de les llengües del món gairebé no tenen representació en línia (Choudhury, 2008). El 2012, META-NET va publicar un estudi en què afirmava que vint-i-una llengües europees, incloent-hi la catalana, estaven en perill d'extinció (Uszkoreit i Rehm, 2012). Woodbury (2019) considera que l'any

2100 el 90 % de les més de set mil llengües parlades al món poden haver desaparegut.

En part, això és degut al fet que el nucli de la tecnologia d'IA actual es troba en poder d'unes quantes empreses privades que segueixen interessos comercials i s'esforcen poc per desenvolupar un producte per a una llengua minoritzada. Si bé és cert que algunes grans corporacions poden incorporar llengües minoritzades a les seves tecnologies, moltes ho fan amb finalitats simbòliques; practiquen així el *language diversity washing*. Aquesta estratègia no aborda genuïnament els problemes subjacents de la preservació, la representació i l'equitat de les llengües i, per tant, els esforços en aquest sentit no sembla que siguin la solució al problema.

Una de les raons per les quals les corporacions no invertiran en el desenvolupament de productes per a llengües amb menys recursos és la manca de dades, que és el primer pas per poder construir qualsevol sistema d'IA. En el cas de la tecnologia del llenguatge, amb dades ens referim a les grans quantitats de text o d'enregistraments de veu que fan falta per desenvolupar un model de NMT o un assistent de veu. En el context de la preservació d'una llengua en línia, les dades esdevenen un bé comú i és aquí on les organitzacions institucionals entren en joc. Els governs poden opinar sobre aquest assumpte i poden desenvolupar programes d'IA específics per a tecnologies lingüístiques que assegurin que hi ha la infraestructura lingüística necessària en termes de dades i models. Així ha estat en el cas de l'irlandès (Ní Chasaide et al., 2019) i el gal·lès (Prys et al., 2019), i més recentment en el del basc (GAITU), el gallec (Proxecto NÓS) i el català (Projecte AINA). Malauradament, aquest enfocament depèn del poder institucional i no és factible per a totes les

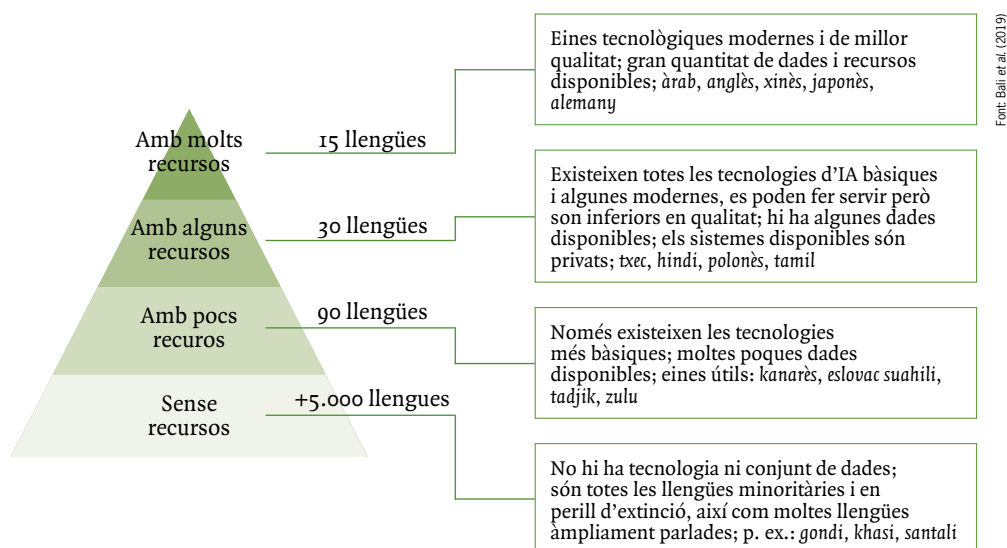


FIGURA 1. Classificació de les llengües segons la disponibilitat de tecnologia, d'eines i de recursos lingüístics

llengües. Tot i així, l'impuls creixent de la IA i la tecnologia del llenguatge pot encoratjar altres governs a desenvolupar plans similars.

En aquest panorama més aviat ombrívol és on trobem una esclatxa d'esperança per a la digitalització de la llengua, que obriria noves oportunitats per a la conservació de les llengües i la superació de les barreres lingüístiques. Per primera vegada, el destí d'una llengua està lligat a la seva comunitat, independentment del nombre de parlants. En els darrers anys han sorgit múltiples iniciatives col·laboratives autogestionades, que han suposat una transformació radical dels seus llenguatges respectius. Amb aquesta nova perspectiva, la comunitat s'implica no només com a consumidora de tecnologia, sinó també com a productora de nous materials i recursos (Prys et al., 2019).

Amb més de 10.048.969 parlants (Plataforma per la Llengua, 2018), el català ocupa el lloc 127 en la classificació de llengües d'Ethnologue segons el nombre de parlants (Eberhard et al., 2023). Així mateix, hi ha més de 110.000 dominis «.cat» (Fundació.cat, 2022) i el català és la desena llengua europea més activa a Twitter i la dinovena a escala mundial (Plataforma per la Llengua, 2020). És clar, doncs, que la seva representació en el món en línia és molt més gran que la que tindria per nombre de parlants i estatus sociopolític (Irigoyen et al., 2020; Melero, 2021); per tant, en comparació amb altres llengües minoritàries, el català té un alt potencial de mobilització de la comunitat (Külebi, 2021).

Una de les principals forces de promoció del català a la Xarxa és Softcatalà, una associació sense ànim de lucre fundada el 1997 amb l'objectiu de promoure l'ús del català en l'àmbit de les tecnologies de la informació (TI). Són els desenvolupadors d'una àmplia gamma d'eines de codi obert, d'entre els quals, els més coneguts són un corrector ortogràfic gramatical i un traductor. També van ser els desenvolupadors del primer assistent de veu en català, anomenat Ona, que es va basar en Catrotron, el primer sintetitzador de veu català construït per Col·lectivaT, una altra associació sense ànim de lucre que desenvolupa tecnologies lingüístiques per al català. Tots són projectes de codi obert i, per tant, representen un exemple clar de com la comunitat es beneficia de la comunitat.

Softcatalà també s'ha implicat en la promoció del projecte Common Voice, un projecte col·laboratiu iniciat per Mozilla el 2017 que té com a objectiu recopilar grans quantitats de dades de veu oberta per crear sistemes de reconeixement automàtic. Actualment, el projecte AINA també hi ha participat i ha portat el català a ocupar la posició número dos pel que fa a hores de gravació recollides en el projecte Common Voice, només cent hores per darrere de l'anglès. Altres exemples reeixits del projecte Common Voice per a llengües amb menys recursos han tingut lloc a Islàndia (Mollberg et al., 2020) i Ruanda (Muhire, 2020).

Un altre projecte internacional i imprescindible en què el català té una gran participació és la Viquipèdia; l'enciclopèdia en línia, oberta i gratuïta. La comunitat catalana a la Viquipèdia és molt activa i ha situat la versió en català en la posició vint en nombre d'articles a escala mundial, de les 333 Viquipèdies que existeixen, amb més de 725.000 articles. En termes de qualitat, la Viquipèdia en català també lidera el rànquing de qualitat dels 1.000 articles que tota Viquipèdia hauria de tenir (Hinojo, 2020). La importància de la Viquipèdia rau en el fet que permet la reutilització dels seus continguts i això la converteix en un recurs molt valuós per al desenvolupament de la tecnologia del llenguatge.

Aquests són només alguns exemples, però ja podem veure que totes les iniciatives no només es caracteritzen pel voluntariat i la col·laboració de la societat civil, sinó també pel fet que tots els recursos i sistemes generats són de codi obert i amb llicències permissives que en permetin la reutilització.

Una altra iniciativa que segueix la ciència oberta va néixer el 2022 des de l'acadèmia anomenada BigScience. Com a resposta al ràpid creixement dels LLM de propietat privada, més de nou-cents investigadors de tot el món van unir forces per desenvolupar el primer model massiu generatiu i en obert, BLOOM (Scao et al., 2020). Curiosament, les llengües d'aquest model inclouen el català, el basc i el nigerocongolès, mentre que hi falten algunes de les llengües «grans» (vegeu la figura 2). Això va ser perquè qualsevol persona interessada podia participar-hi i incloure-hi el seu idioma. Les comunitats actives darrere dels idiomes esmentats anteriorment es van assegurar que els seus hi fossin presents i en van contribuir així a la preservació, tot aprofitant el poder del col·lectiu.

En conclusió, hem vist per què la indústria de la intel·ligència artificial es concentra en gran manera en un grapat de llengües franques i com una comunitat activa pot ser el factor clau per a la preservació de les llengües en els temps actuals. La generació de nous recursos, la implicació constant de la comunitat, així com la creació de les infraestructures públiques necessàries, poden alterar el destí d'una llengua en perill d'extinció digital i convertir-la en una llengua en expansió digital.

És clar que «sense comunitat no hi ha dades» (Muhire, 2020) i, per tant, no hi ha IA. No obstant això, en el cas concret del català, podem capgirar la citació i afirmar que, si hi ha una comunitat, hi ha dades i, amb dades, el desenvolupament de la tecnologia lingüística del català acaba de començar. De fet, d'acord amb l'últim informe del European Language Equality Project (Melero et al., 2022), el català és la llengua d'àmbit regional que té el suport més gran en tecnologia lingüística (vegeu la figura 3), seguida del basc, el gal·lès i el gal·lès; totes són llengües que reben suport institucional.

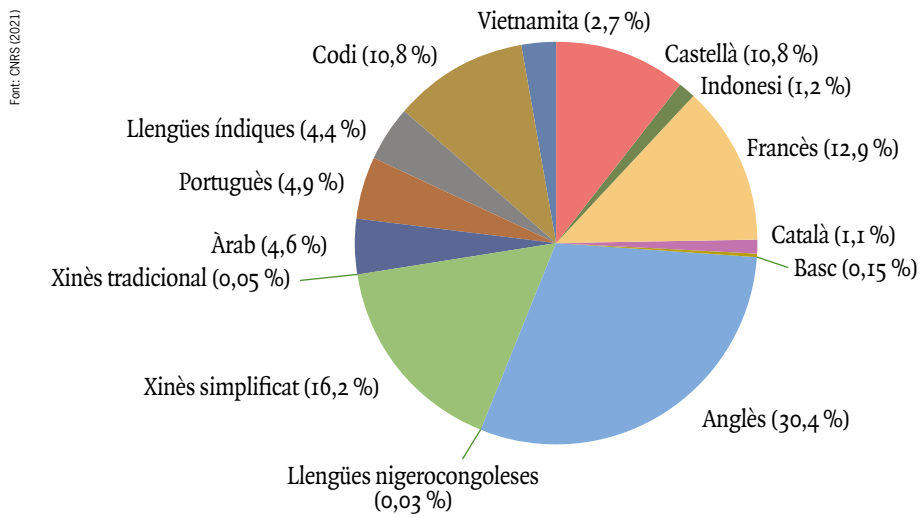


FIGURA 2. Llengües utilitzades per entrenar BLOOM

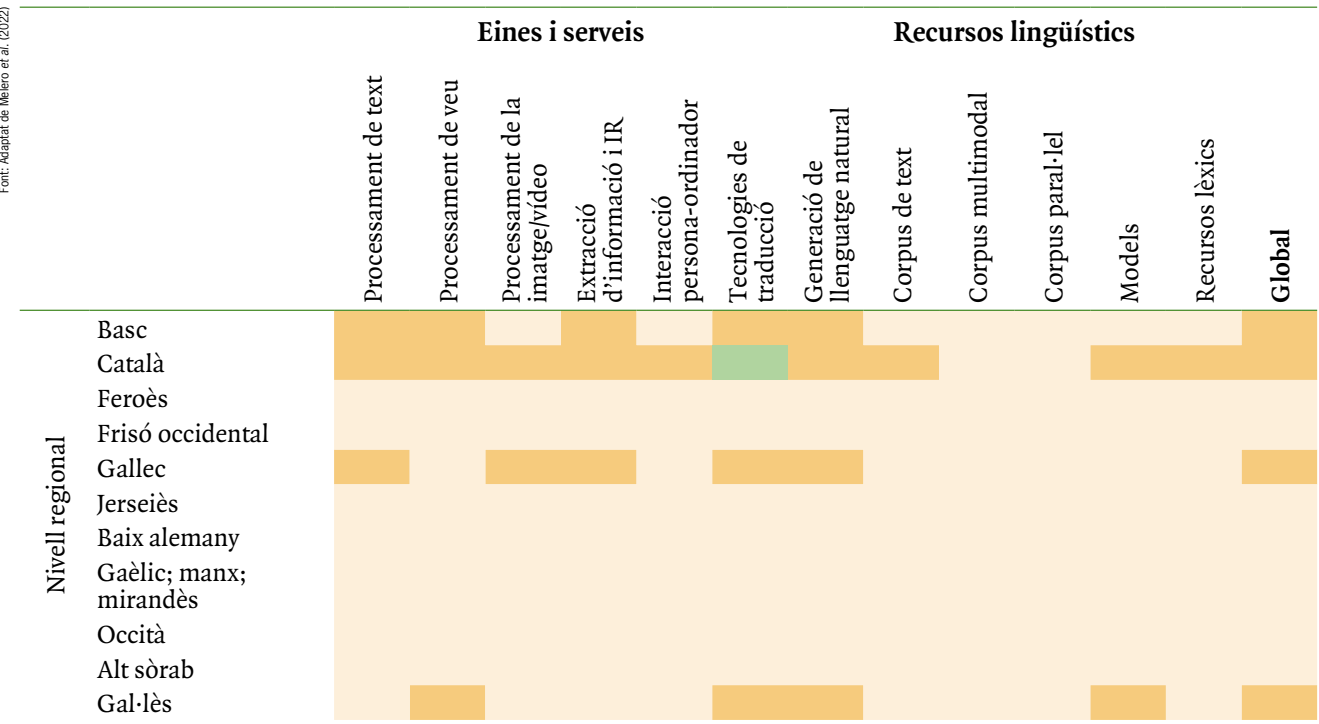


FIGURA 3. Estat de suport tecnològic, el 2022, de les llengües europees regionals, pel que fa a les eines i serveis i als recursos lingüístics (groc clar: suport feble o gens de suport; groc: suport fragmentari; verd clar: suport moderat)

La distribució entre llengües amb molts recursos i amb menys recursos és un espectre que canvia contínuament. Fa uns quants anys, el català era una llengua de recursos escassos, però gràcies a l'activa comunitat en línia que ha establert les bases de

la tecnologia lingüística, ja no és així. Per tant, si vols contribuir a preservar la teva llengua, «només» necessites escriure, parlar i fer servir les xarxes socials en català. D'aquesta manera, seràs un activista de la llengua, digitalment resiliència. 🌱

Bibliografia

- BALI, Kalika; CHOUDHURY, Monojit; SITARAM, Sunaya; Seshadri, Vivek (2019). «ELLORA: Enabling low resource languages with technology». *Proceedings of the 1st International Conference on Language Technologies for All*, p. 160-163.
- CHOUDHURY, Monojit (2008). «Breaking the Zipfian Barrier of NLP». *Proceedings of the IJCNLP- 08 Workshop on NLP for Less Privileged Languages*.
- CNRS (2021). *Largest trained open-science multilingual language model ever* [en línia]. Comunicat de premsa. <<https://www.cnrs.fr/en/release-largest-trained-open-science-multilingual-language-model-ever>> [Consulta: 29 de març de 2023].
- EBERHARD, David M.; SIMONS, Gary F.; FENNIG, Charles D. (ed.) (2023). *Ethnologue: Languages of the world* [en línia]. 26a ed. Dallas, Texas: SIL International. <<http://www.ethnologue.com>> [Consulta: 29 de març de 2023].
- FUNDACIÓ.CAT (2022). *Estat del català a Internet i les TIC* [en línia]. <<https://observatori.fundacio.cat/#evolucio>> [Consulta: 29 de març de 2023].
- HINOJO, Àlex (2020). «Somien els viquipedistes en enciclopèdies elèctriques? Present i futur de la Viquipèdia i el rol de la comunitat catalanoparlant». *Revista de Llengua i Dret / Journal of Language and Law* [en línia], 73, 133-145. <<https://doi.org/10.2436/rld.i73.2020.3424>> [Consulta: 29 de març de 2023].
- KÜLEBI, Baybars (2021). «El fet diferencial del català: la comunitat de programari lliure i obert». *Pensem* [en línia] <<https://www.pensem.cat/noticia/226/fet-diferencial-catala-la-comunitat-programari-lliure-obert>> [Consulta: 29 de març de 2023].
- MELERO, Maite (2021). «La tecnologia obre portes a la diversitat lingüística digital». *Pensem* [en línia]. <<https://www.pensem.cat/noticia/233/maite-melero--tecnologia-obre-portes-diversitat-linguistica-digital>> [Consulta: 29 de març de 2023].
- MELERO, Maite; FIGUERAS, Blanca C.; RODRÍGUEZ, Mar; VILLEGAS, Marta (2022). «Report on the Catalan language». *Language Technology Support of Europe's Languages in 2020/2021 - European Language Equality Project*.
- MOLLBERG, David Erik; JÓNSSON, Ólafur. Helgi; PORSTEINSDÓTTIR, Sunneva; STEINGRÍMSSON, Steinnþór; MAGNÚSDÓTTIR, Eydis Huld; GUÐNASON, Jón (2020). «Samrómur: Crowd-sourcing data collection for Icelandic speech recognition». *LREC 2020 - 12th International Conference on Language Resources and Evaluation, Conference Proceedings* (maig), p. 3463-3467.
- MUHIRE, Remy (2020). *How Rwanda is making voice tech more open* [en línia]. Mozilla Foundation. <<https://foundation.mozilla.org/en/blog/how-rwanda-making-voice-tech-more-open/>> [Consulta: 29 de març de 2023].
- NÍ CHASAIDE, Ailbhe; NÍ CHIARÁIN, Neasa; BERTHESEN, Harald; WENDLER, Chrisyoph; MURPHY, Andrew; BARNES, Emily; GOBL, Christer (2019). «Can we defuse the digital timebomb? Linguistics, speech technology and the Irish language community». *Proceedings of the 1st International Conference on Language Technologies for All* [en línia], p. 177-181. <<https://lt4all.elra.info/proceedings/lt4all2019/pdf/2019.lt4all-1.45.pdf>> [Consulta: 29 de març de 2023].
- PLATAFORMA PER LA LLENGUA (2018). *InformeCAT 2018: 50 dades sobre la llengua catalana* [en línia]. <https://www.plataforma-llengua.cat/media/upload/pdf/informecat2018_1528713023.pdf> [Consulta: 29 de març de 2023].
- (2020). *InformeCAT 2020: 50 dades sobre la llengua catalana* [en línia]. <https://www.plataforma-llengua.cat/media/upload/pdf/informecat-2020_267_11_2406.pdf> [Consulta: 29 de març de 2023].
- PRYS, Delith; JONES, Dewi B.; PRYS, Gruffud (2019). «Planning for language technology development and language revitalization in Wales». *Proceedings of the 1st International Conference on Language Technologies for All*, p. 367-370.
- RIERA IRIGOYEN, Marc; IVERS RIBES, Xavier; ORGA ESTEVE, Pere; MONTANÉ CAMACHO, Joan, MAS HERNÁNDEZ, Jordi; VICEDO CREMADES, Artur (2020). «Softcatalà: nous reptes per garantir la vitalitat del català a les tecnologies». *Revista de Llengua i Dret / Journal of Language and Law* [en línia], 73, p. 146-153. <<https://doi.org/10.2436/rld.i73.2020.3396>>. [Consulta: 29 de març de 2023].
- SCAO, Teven Le; FAN, Angela; AKIKI, Christopher; PAVLICK, Ellie; ILIĆ, Suzana; HESSLOW, Daniel; [...] WOLF, THOMAS. (2022). «Bloom: A 176b-parameter open-access multilingual language model» [en línia]. <<https://arxiv.org/abs/2211.05100>>. [Consulta: 29 de març de 2023].
- USZKOREIT, Hans; REHM, Georg (2012). *META-NET White Paper Series: Press Release* [en línia]. <<http://www.meta-net.eu/whitepapers/press-release>>. [Consulta: 29 de març de 2023].
- WOODBURY, Anthony C. (2019). *What is an endangered language?* Washington: Linguistic Society of America.

Crònica de la XX Jornada de la SCATERM

XX Jornada de la SCATERM

«La terminologia en l'educació: entre la innovació i l'abús»

31 de maig de 2023

Sala Pere i Joan Coromines de l'IEC i virtual



Font: SCATERM

Participants de la primera taula rodona: Neus Doñate, Rafel Lemus, Neus Lorenzo, Josep Serentill i Teresa Tort

El dimecres 31 de maig a la tarda va tenir lloc, en format presencial i virtual, la XX Jornada de la SCATERM, amb el títol «La terminologia en l'educació: entre la innovació i l'abús». La jornada es plantejava analitzar la creació constant de terminologia pedagògica i la seva comprensió per part de la comunitat educativa, així com valorar l'ús de terminologia dins de diferents àrees de coneixement en l'educació, preguntant-se, per exemple, quina i quanta terminologia s'ha d'usar en els diferents nivells educatius. Per tal de tenir una visió polièdrica del tema, vam comptar amb intervencions de la Societat Catalana de Pedagogia, de serveis educatius, d'especialistes d'universitats i del món editorial.

La presidenta de la SCATERM, Ester Bonet, i el president de la Secció Filològica de l'IEC, Nicolau Dols, van fer els parlaments inicials.

A continuació van tenir lloc dues taules rodones. La primera, moderada per Teresa Tort, portava per nom «**El desgast terminològic en l'àmbit pedagògic**» i s'hi van analitzar els efectes dels canvis constants en la terminologia d'aquest sector. Josep Serentill, president de la Societat Catalana de Pedagogia, remarca que la llengua està en la base de tots els aprenentatges escolars i que, a més, és l'eina principal de cohesió social. Neus Lorenzo, també de la Societat Catalana de Pedagogia, va mostrar-nos les respostes «políticament correctes» de la nova tec-

nologia del ChatGPT a la pregunta sobre l'abús terminològic en l'àmbit educatiu. També ens recordà que l'ús de certa terminologia dona control a les persones que la utilitzen. Rafel Lemus Arce, formador i docent de secundària a Catalunya, parlà sobre la funció de la terminologia en les àrees del coneixement a partir de la seva experiència professional. També mostrà com les diferents lleis educatives, de 1970 a 2020, han anat variant la terminologia, es preguntà si en altres àmbits socials també hi ha tanta quantitat d'instruccions i mostrà, així, la necessitat d'una formació bàsica per al professorat. Raquel Doñate Calpe, assessora del Servei d'Educació Plurilingüe del País Valencià, va fer refe-



Font: SCATERM

Participants de la segona taula rodona: Rosa Estopà, Núria Planas, Alícia Almonacid i Francesc Galera

rència a l'alumnat i les famílies com a part integrant dels canvis terminològics en l'educació es va plantejar alguns dels objectius d'aquests canvis, com poden ser homogeneïtzar, eliminar doblats o reflectir l'enfocament inclusiu. Aquesta primera taula rodona acabà amb un torn de debat entre els ponents.

Després de la pausa del cafè, començà la segona taula rodona, moderada per Francesc Galera, amb el títol «**Quina i quanta terminologia s'ha d'usar en els diferents nivells educatius?**». Rosa Estopà, directora del Vocabulari especialitzat bàsic (VEB), ens mostrà com el lèxic especialitzat comença en els primers anys de vida. També ens explicà les fases de treball del VEB. Estopà creu que és necessari conèixer la llista de termes

imprescindibles en cada nivell escolar i disposar de diccionaris especialitzats per a cada nivell educatiu. Finalment, ens presentà el Glossari de la Gramàtica bàsica de l'IEC que realitzà la professora Cuenca. Núria Planas, professora de didàctica de les matemàtiques de la Universitat Autònoma de Barcelona, parlà dels problemes d'ús en alguns termes matemàtics i va incidir en la diferència que hi ha entre saber la paraula (naming) i saber-la explicar (explaining), cosa que pot provocar desavantatges lingüístics. Alícia Almonacid, coordinadora de l'àmbit científic de l'editorial Barcanova, parlà de la difícil connexió que hi ha entre els currículums i els projectes editorials, així com del lèxic específic que arriba a l'alumnat. Finalment, apel·là

a poder despertar l'interès i l'esperit crític dels alumnes. Amb un torn de debat acabà aquesta segona taula rodona, en què es parlà de les relacions entre la política i les terminologies de l'educació.

Ester Bonet va cloure la sessió, recordant el concepte d'estrès terminològic que havien utilitzat alguns ponents i com les lleis són font d'una innovació terminològica a la qual ens hem d'adaptar, sense oblidar que els docents són els qui han de donar sentit a la paraula. Finalment, va agrair l'assistència al públic, presencial i virtual, i a tots els ponents d'aquesta vintena jornada.

JOAN PUIGMALET
Societat Catalana de Terminologia

Les Tertúlies Terminològiques de RITerm: un espai de debat sobre terminologia



L’acta de constitució de la Xarxa Iberoamericana de Terminologia (RITerm) indica que «l’intercanvi d’informació terminològica, la cooperació en la posada en marxa i el desenvolupament de projectes relacionats amb la terminologia i la cooperació en la formació en terminologia dels participants» són les activitats centrals de la xarxa. Aquesta missió ha estat, des de 1988 fins avui, el motor de moltes iniciatives que s’hi han organitzat amb la finalitat de desenvolupar i difondre la terminologia.

El Simposi Iberoamericà de Terminologia, que se celebra cada dos anys, és una mostra d’aquesta mena d’activitats; tanmateix, vam creure que aquest espai no era suficient i que ens feia falta tenir més trobades per tal d’obrir el diàleg entre col·legues que dia a dia promouen projectes relacionats amb l’àmbit de la terminologia. Trobar-nos cada dos anys era poc per divulgar els projectes que sorgeixen dins de la família RITerm, i per això vam construir l’espai denominat Tertúlies Terminològiques (<https://bit.ly/RITermYouTube>), el propòsit del qual és donar a conèixer projectes de ter-

minologia iberoamericana, consolidar llaços entre diversos grups de recerca i associacions, i crear entorns de debat sobre la terminologia com a disciplina.

El juny de 2022 vam començar aquesta aventura terminològica convidant la Societat Catalana de Terminologia a la tertúlia amb el títol «SCATERM: punto de encuentro de la terminología catalana», a càrrec d’Ester Bonet i Solé i de Mercè Lorente i Casafont, que van exposar la història de la societat, des que va fundar-se com a Associació Catalana de Terminologia (2002) fins que va esdevenir societat filial de l’Institut d’Estudis Catalans amb el nom de SCATERM (2008). Van referir-se als seus objectius: ser el punt de trobada de la terminologia catalana, difondre la terminologia en llengua catalana en àmbits científics i tècnics, promoure l’intercanvi d’informació sobre activitats terminològiques i establir relacions amb altres organitzacions. A més, van presentar la seva revista *Terminàlia*, i van comentar diversos esdeveniments organitzats i diverses obres de divulgació de la terminologia

en llengua catalana que s’havien publicat.

El setembre de 2022 vam seguir el nostre viatge, aquest cop pel Brasil, amb una tertúlia titulada «A pesquisa terminológica no grupo Termisul», a càrrec de Cleci Bevilacqua i Sandra Loguercio, que ens van explicar com es va formar el grup de recerca TERMISUL a la Universitat Nacional de Rio Grande do Sul (UNRGS) l’any 1999 i ens van presentar la seva història a partir de la producció de recerques teòriques i aplicades que enriqueixen el panorama de la terminologia en la llengua portuguesa del Brasil.

Més endavant, el novembre de 2022, vam viatjar a Mèxic amb la participació en la tertúlia de Gerardo Sierra, que ens va parlar sobre «GECO, un servicio web para la terminología». El professor Sierra hi va presentar diverses eines per a l’elaboració de corpus i es va centrar en el gestor de corpus GECO, desenvolupat pel Grupo de Ingeniería Lingüística (GIL) de la Universidad Nacional Autónoma de México, com va néixer, com funciona la nova versió i quina feina de futur queda encara per fer.

El febrer de 2023 vam arribar a Colòmbia, amb una tertúlia sobre un «Breve recorrido de los trabajos de terminología del Grupo TNT», en què es va fer una presentació preliminar del *Nuevo manual de terminología del español*, a càrrec de Gabriel Quiroz, Diego Burgos i Felipe Zuluaga. En aquesta trobada van exposar la història del Grupo TNT i la seva producció acadèmica i van aprofundir en la seva obra més recent: un manual de terminologia i els reptes amb què s'enfrontaren per realitzar aquest producte.

L'abril de 2023 ens vam traslladar fins al Con Sud, on Paola Cañete-González i Constanza Becerra ens van presentar «NeoUdeC: investigando la contingencia neológica en el español de Chile», un grup que fa recerca sobre la neologia de l'espanyol de Xile. En aquesta trobada van comentar com recopilen i classifiquen els neologismes, com seleccionen el neologisme del mes i com conviden a participar la societat en tot aquest procés, perquè la comunicació cal que sigui extramurs, és a dir, fora dels murs de la universitat.

Totes aquestes tertúlies ens acosten i enriqueixen el nostre espai de diàleg, promouen la missió de la xarxa RITerm i demostren el dinamisme i l'entusiasme de la recerca terminològica a Iberoamèrica.

Enllaç a les tertúlies: <https://bit.ly/RITermYouTube>.

IOANA CORNEA
Comitè Executiu de RITerm
Universidad Autónoma de México

Traducció del castellà a càrrec de Mercè Lorente
(Consell de Redacció)

El refrany

Paraula és dada a tots, mes saviesa a pocs

Font: Paremiologia catalana. www.dites.cat

Recull de les activitats de la Societat Catalana de Terminologia

L'Associació Catalana de Terminologia (ACATERM), creada l'any 2002, va organitzar la seva primera trobada pública, amb el nom de I Jornada de Terminologia i Traducció, que es va celebrar a la Universitat Jaume I (UJI) de Castelló de la Plana el 9 de maig de 2003. La Jornada va ser organitzada conjuntament pel Servei de Llengües i Terminologia i el Departament de Traducció i Comunicació de l'UJI, l'Institut Universitari Joan Lluís Vives (en endavant, Xarxa Vives

d'Universitats) i l'ACATERM, institució que agafaria el relleu de l'IULA en la celebració de trobades monogràfiques sobre terminologia i alguna matèria afí.

A partir de l'any 2008, l'ACATERM va ser reconeguda com a societat filial adscrita a la Secció Filològica de l'Institut d'Estudis Catalans (IEC) i va passar a anomenar-se Societat Catalana de Terminologia. Formar part de l'IEC va suposar l'inici de la col·lecció «Memòries de la Societat Catalana de Terminologia»,

que aplega les actes de les activitats d'aquesta Societat a partir de l'any 2009, que més endavant, l'any 2015, van passar a ser gravades i pujades al canal de YouTube de l'IEC. Tota aquesta informació està disponible en accés obert a Internet a través de la plataforma IEC Obert (<https://iecobert.iec.cat/>).

Seguidament fem un repàs de les activitats dutes a terme al llarg d'aquests vint anys:



V Seminari, de 2010, a la Reial Acadèmia de Medicina

Recull de les activitats de la Societat Catalana de Terminologia

Societat Catalana de Terminologia



VII Seminari, de 2012, a la Facultat de Química



Taula rodona 2014, «La terminologia instrumentalitzada»

Jornades de l'ACATERM		Any
I «Terminologia i traducció»		2003
II «Terminologia i estandardització científica»		2004
III «Terminologia i consum: etiquetatge i publicitat de productes alimentaris»		2005
IV «Terminologia i traducció: traducció de productes informàtics»		2006
V «Terminologia i serveis lingüístics»		2007
VI «Terminologia de la gastronomia i de l'enologia»		2008

Jornades de la SCATERM		
VII «Terminologia i documentació»		2009
VIII «Terminologia, neologia i traducció»		2010
IX «Terminologia i llenguatge jurídic»		2011
X «Els manlleus en la terminologia musical»		2012
XI «La terminologia en el món educatiu: didàctica, recursos i estratègies»		2013
XII «Presència del català en l'espai científicotècnic. Quina terminologia tenim, sense textos especialitzats?»		2015
XIII «La terminologia abans de la Terminologia: del segle XIII al XVII»		2016
XIV «La terminologia de la Gramàtica»		2017
XV «La terminologia de l'esport»		2018
XVI «La Viquipèdia i la terminologia»		2019
XVII «La terminologia del diccionari»		2020
XVIII «La terminologia en la comunicació de la COVID-19»		2021
XIX «La terminologia a peu de carrer»		2022
XX «La terminologia en l'educació: entre la innovació i l'abús»		2023

Taules rodones		
«El concepte de terme»		2011
«La terminologia instrumentalitzada. Quan els termes no són neutres ni innocents»		2014
«La IATE: un recurs útil per a la traducció?»		2019

Seminars de terminologia		
I «Nou marc teòric i conseqüències en el treball terminològic»		2006
II «Tecnologia per al treball en terminologia»		2007
III «El programa Terminus, una eina de gestió»		2009
V «L'establiment de paraules clau i la redacció d'índexs de llibre»		2010
VI «Terminologia de les ciències de la vida: criteris i recursos»		2011
VII «La nomenclatura química de la IUPAC en català»		2012

Curs		
I Curs de Correcció Avançada per a l'Edició de textos Científics i Tècnics		2010

SOCIETAT CATALANA DE TERMINOLOGIA

Juan Carlos Sager (1929-2021)



Juan Carlos Sager: home polifacètic, avançat al seu temps

MARÍA POZZI

Centro de Estudios Lingüísticos y Literarios
El Colegio de México

GERARDO SIERRA

Grupo de Ingeniería Lingüística
Instituto de Ingeniería
Universidad Nacional Autónoma de México

TERMINÀLIA 27 (2023): 97-100 · DOI: 10.2436/20.2503.01.192

ISSN: 2013-6692 (impresa); 2013-6706 (electrònica)

<http://terminalia.iec.cat>

Juan Carlos Sager deixa un gran llegat no només de publicacions, sinó també de molts altres elements, igual d'importants. I els records d'aquests perduren en la memòria de tothom que el va conèixer i admirar. La majoria de persones que han abordat professionalment els llenguatges d'especialitat, la traducció, la lexicografia, la traducció automàtica, l'enginyeria lingüística, la tecnologia de la informació i, sobretot, la terminologia, han conegut en Juan Carlos a través de les seves paraules. Però els que vam tenir la fortuna de conèixer-lo personalment —perquè va fer-nos de professor o n'érem col·legues, companys o amics— el recordem, a més, com algú que, tot i que inicialment imposava i podia ser una mica distant, sempre va compartir desinteressadament els seus coneixements i les seves experiències acadèmiques, i amb qui vam establir una relació de confiança i respecte mutus.

El professor Sager va néixer a l'Argentina el 3 d'agost de 1929. Va iniciar els estudis universitaris a Tucumán. Tot i que era molt jove, ja tenia una curiositat i un interès científic molt grans pels temes més capdavaners en lingüística aplicada. Després, va anar als Estats Units, a la Universitat de Notre Dame, i més tard a la Gran Bretanya, a Manchester, on va romandre la resta de la seva vida i on es van gestar les seves aportacions més destacades.



Font: arxiu personal

Juan Carlos Sager: home polifacètic, avançat al seu temps

MARÍA POZZI I GERARDO SIERRA



Juan Carlos Sager va ser un home polifacètic, un pensador original, un professor generós i compromès, un escriptor prolífic, un innovador, un acadèmic visionari i un pioner en els camps de la traducció, la terminologia i la lingüística computacional. Era un avançat a l'època en què va viure.

En els àmbits acadèmic i de gestió universitària, en Juan es va distingir per les seves iniciatives creatives. Es va implicar activament en la creació del Departament de Llengües Modernes de la Universitat de Salford. El 1980, el van nomenar cap del Departament d'Estudis Europeus i de Llengües Modernes de la Universitat de Manchester - Institute of Science and Technology (UMIST). Va impulsar-hi canvis que van modernitzar el Departament i van donar-hi un gir avantguardista, tot introduint-hi les darreres novetats en informàtica per optimitzar el treball en lingüística aplicada, especialment en els camps de la traducció, la lexicografia i la terminologia. Va tenir un paper cabdal en la creació del Centre for Computational Linguistics (CCL), que tenia com a objectius principals generar recerca capdavantera en lingüística computacional i formar-hi estudiants d'arreu del món i en tots els nivells (des del grau fins al doctorat). En conseqüència, nombrosos països encara ara acullen investigadors i lingüistes computacionals amb carreres molt reeixides que van estudiar al CCL/UMIST.

En Juan fou ben reconegut en nombrosos àmbits europeus, atesa la seva capacitat de treballar i d'obtenir resultats, així com pels seus dots organitzatius. Per tant, és natural que s'hi atansessin per col·laborar-hi

en projectes com EUOTRA, de traducció automàtica entre les llengües de la Comunitat Econòmica Europea (precedents de l'actual Unió Europea), dut a terme entre 1978 i 1992. Sager sempre estava predisposat a explorar temes de recerca innovadors, de manera que el CCL/UMIST també va assolir l'èxit amb altres projectes europeus sobre lingüística en terminologia i lexicografia multilingües.

Alguns dels llibres que va publicar han esdevingut obres de referència dels temes de què s'ocupava, tant des d'un vessant acadèmic com pel que fa a la pràctica professional. La seva obra més influent en terminologia és *A practical course in terminology processing*, de 1990. Sager hi aborda l'estudi de la terminologia des de tres dimensions: la cognitiva, la lingüística i la comunicativa. La tesi de l'obra posa en entredit diversos fonaments exposats a la teoria general de la terminologia (TGT), proposada per Eugen Wüster i que Helmut Felber va publicar el 1979 a títol pòstum. Cal tenir en compte que en aquella època la TGT dominava bona part del treball fet i publicat fins aleshores en terminologia teòrica i aplicada. El *Course* també inclou alguns apartats sobre estandardització i gestió terminològica per a l'emmagatzematge, la recuperació i l'ús de la terminologia.

Cal recordar que Teresa Cabré va publicar l'any 1992 en català la seva primera gran obra en terminologia, *La terminologia. La teoria, els mètodes, les aplicacions*. Després, se'n van publicar les traduccions al castellà (1993) i a l'anglès (1999), i Sager va encarregar-se d'aquesta darrera. L'obra de Cabré coincideix amb Sager en alguns aspectes, com ara l'aproximació des d'aquestes tres dimensions cognitiva, lingüística i comunicativa. En destacar la dimensió comunicativa de l'essència i la raó de ser de la terminologia, Cabré va obrir la porta a noves posicions teòriques i metodològiques que predominen en terminologia avui, en particular, la teoria comunicativa de la terminologia (TCT) i les idees que n'han sorgit a posteriori. Des d'aleshores, una gran amistat va unir en Juan i na Teresa, un lligam de molts anys, que durà fins a la mort d'en Juan. Sens dubte, aquesta unió ha beneficiat molts terminòlegs.

Abans del *Course*, Sager va publicar, juntament amb David Dungworth i Peter McDonald, el volum *English special languages. Principles and practice in science and technology* (1980). Els autors hi presenten una anàlisi, una caracterització i una classificació detallada dels llenguatges d'especialitat. Més tard, va publicar dos llibres més: *Language engineering and translation. Consequences of automation* (1994), el qual descriu el paper de la tecnologia de la informació com a eina per a la traducció, i *Essays on definition* (2000), volum en què examina la teoria de la definició des de Plató i Aristòtil fins a l'època moderna des d'una perspectiva filosòfica i en el qual demostra la importància de les relacions semàntiques en el context de la terminologia.

En l'àmbit editorial, Sager també va deixar un llegat important. Va fundar tres revistes científiques de gran prestigi (o va ser-ne redactor en cap), cadascuna de les quals s'ha consolidat com un espai de debat internacional dels àmbits respectius. La Comissió de les Comunitats Europees (actualment, la Unió Europea) el va convidar a ser el primer editor en cap de *Multilingua*, una revista que tracta de temes de comunicació relacionats amb la diversitat cultural i lingüística, i explora com es podria aplicar la tecnologia lingüística per resoldre, almenys parcialment, els problemes associats a aquestes noves realitats. El primer número va aparèixer l'any 1982 a Mouton Publishers (ara, De Gruyter Mouton). La segona revista, la fundació de la qual va promoure sense defallir, va ser *Meta* (Presses de l'Université de Montréal). Aquesta revista està especialitzada en traducció, interpretació, estudis de traducció, estudis de terminologia comparada, traducció assistida per ordinador i documentació. Per acabar, però no per això menys important, Sager va contribuir a endegar-ne una tercera: *Terminology* (John Benjamins), una revista que aborda problemes lingüístics de la traducció (com l'ambigüitat) i la comunicació interdisciplinària; tracta, a més, de la representació del coneixement, les eines de la tecnologia de la informació i les bases de dades terminològiques. Se'n publicà el primer volum el 1994, al qual Sager va contribuir amb tres articles (primer volum, números 1 i 2): «Terminology: Custodian of knowledge and means of knowledge transfer», «What's wrong with "terminology work" and "terminology science"?» i, juntament amb Marie-Claude L'Homme, «A model for the definition of concepts: Rules for analytical definitions in terminological databases». Entre 1994 i 2015 va ser editor assessor de *Terminology* i de la *Language Companion Series* (John Benjamins).

Per si no n'hi hagués prou, J. C. Sager també va assessorar diversos organismes nacionals i internacionals. Entre d'altres, hi ha la Unió Europea, la UNESCO, l'INFOTERM o la British Standards Institution (BSI), organisme britànic membre de l'Organització Internacional per a la Normalització (ISO), en el qual va dirigir l'equip britànic de treball de l'ISO/TC 37 (Terminologia: principis i coordinació), que actualment té la denominació ISO/TC 37 (Llengua i terminologia).

No en va, l'any 2001 va rebre el Premi Eugen Wüster per la seva contribució destacada a la terminologia.

En Juan era un ésser humà singular. Tal com tenia interessos acadèmics diversos, la seva vida privada també era polièdrica: gaudia de la companyia de família i amics, viatjava, sortia en bicicleta, nedava, passejava pels jardins Fletcher Moss i s'atansava a peu fins a Didsbury a comprar habitualment. S'apassionava amb l'art, la literatura, la música, la filosofia, el menjar gourmet, els trencaclosques i convidant gent a casa. El seu company sentimental durant molt de temps, Giovanni Pontiero, el complementava a la perfecció: era un

acadèmic destacat, tenia sentit de l'humor i el do de la conversa, i era un cuiner excel·lent. En Giovanni traduïa Saramago a l'anglès i va guanyar diversos premis europeus per la seva tasca. L'últim, de fet, el va rebre només unes setmanes abans de morir, el dia del seu seixanta-dosè aniversari, el 10 de febrer de 1996.

1, The Grove, Didsbury... Qui podrà oblidar mai els dinars o sopars als quals vam tenir el plaer d'assistir? Arribàvem tots puntuals, no hi romaníem en excés i habitualment marxàvem tots plegats. La taula, parada amb gran exquisidesa, s'engalanava amb tot de fineses. I el menjar... Quin menjar! En Giovanni sempre preparava un plat diferent i, des d'aquella cuina minúscula, dissenyava un menú farcit de menges refinades. Tothom gaudia escoltant les històries d'en Giovanni hores i hores, segurament ornamentades, de les moltes aventures que havien viscut arreu del món. En Juan, en canvi, acostumava a ser molt callat; de tant en tant, però, s'escapava a la cuina i tornava amb vi o alguna cosa per menjar. Potser només era tímid? L'atabalava el caràcter extravertit d'en Giovanni? De vegades, ho semblava. En ocasions, però, feia comentaris irònics, sarcàstics i directes, molt típics d'ell. Probablement allò era només una cuirassa darrere de la qual s'amagava el veritable Juan Carlos: un home honest i sensible.

Després que en Giovanni morís, la vida d'en Juan es va capgirar. Mai no s'hauria imaginat sobreviure-li durant quasi vint-i-cinc anys exactes. Va haver d'aprendre a viure sol de nou i, malgrat que la seva vida social no es va apagar, certament va minvar. Aquells sopars van tornar-se tardes de te, pastes i galetes; si el temps ho permetia, se celebraven al jardí o al saló d'hivern, els seus dos indrets preferits. Aquests dos racons van testimoniar converses llargues i fructíferes, tant de temes acadèmics com d'altres. Amb més temps lliure, l'any 1997, un any després de la mort d'en Giovanni, en Juan el va homenatjar amb el llibre *The dialogue of the translator*. Giovanni Pontiero, editat per Pilar Orero i ell mateix.

En Juan sempre es va mantenir actiu. Molts de nosaltres, segurament, el recordem al capdamunt d'una escala tot reparant el sostre o arreglant una canonada, cuidant aquell jardí magnífic, fent dissabte o llegint i revisant els manuscrits d'amics, de companys de feina o d'antics deixebles que s'hi adreçaven per tal de conèixer-ne el parer i saber si eren prou bons per a publicar-los.

Molts dels deixebles de Sager que tenien fills petits a l'inici eren reticents a dur-los amb ells a casa del professor, per si s'avorrien o es portaven malament. Però mai no va ser així. Sempre s'hi mostrava afable, els explicava històries captivadores i els ensenyava els darrers trencaclosques i escenografies de cartró que tenia, alguns dels quals els va regalar als fills d'en Gerardo Sierra. Ells els atresoraran per sempre. En resum, els nens estaven encantats d'anar a visitar el professor Sager.

Juan Carlos Sager: home polifacètic, avançat al seu temps

MARÍA POZZI I GERARDO SIERRA

Després d'una vida llarga, interessant i fructífera, Juan Carlos Sager ens va deixar a Manchester el 18 de febrer de 2021, com a conseqüència de l'ictus que havia patit uns dies abans. Poca gent va poder assistir al funeral, per dues raons. Primera: preferia un comiat molt senzill. Segona: s'havien decretat una sèrie de restriccions, a causa de la COVID-19, que restringien les reunions de persones.

Sempre el recordarem per tot allò que va ser: un home únic i un mentor esperonador, avançat a la seva època. Sempre romandrà amb nosaltres i sempre ens meravellarem de com un sol home va ser capaç de transformar tantes vides professionals. ✿



MARÍA POZZI

És professora i investigadora del Centro de Estudios Lingüísticos y Literarios del Colegio de México. Es va doctorar en lingüística computacional a l'Institut de Ciència i Tecnologia de la Universitat de Manchester (UMIST). Les seves àrees d'especialitat són la terminografia, la lexicografia infantil, la normalització

terminològica i la neologia. L'any 2006 va rebre el Wüster Special Prize. Ha col·laborat en més de 30 diccionaris especialitzats i en un nombre semblant de normes ISO; les seves publicacions inclouen més de 100 articles en revistes, capítols de llibres i actes de congressos.



GERARDO SIERRA

És investigador titular de l'Instituto de Ingeniería de la Universidad Nacional Autónoma de México, on va fundar i liderar el Grupo de Ingeniería Lingüística. Va fer el doctorat en lingüística computacional a l'Institut de Ciència i Tecnologia de la Universitat

de Manchester (UMIST). Les seves especialitats són la terminòtica, la lingüística de corpus i l'extracció d'informació. Ha publicat tres llibres i més de 200 articles en revistes, capítols de llibres i actes de congressos.

Publicacions de la Societat Catalana de Terminologia

- * Revista TERMINÀLIA. Té periodicitat semestral i constitueix l'instrument de divulgació científica de la SCATERM.
- * Butlletí de la Societat Catalana de Terminologia. És una publicació bimestral en línia que s'ha editat regularment des del mes de març de l'any 2007 fins al mes de desembre de l'any 2014. S'adreça especialment als socis de la SCATERM i, en general, a totes les persones que estan interessades en la terminologia.
- * Col·lecció «Memòries de la Societat Catalana de Terminologia». Aplega treballs derivats de les activitats més rellevants organitzades per la SCATERM, especialment de les jornades i els seminaris anuals:

FELIU, Judit; TRIAS, Mireia (cur.) (2019). *Gramàtica, esport i terminologia*. Barcelona: Societat Catalana de Terminologia. (Memòries de la Societat Catalana de Terminologia; 7) [Publicació electrònica]

FELIU, Judit; TRIAS, Mireia (cur.) (2021). *Viquipèdia i terminologia*. Barcelona: Societat Catalana de Terminologia. (Memòries de la Societat Catalana de Terminologia; 8) [Publicació electrònica]

TRIAS, Mireia (cur.) (2022). *Terminologia i societat: recursos terminològics i escenaris comunicatius*. Barcelona: Societat Catalana de Terminologia. (Memòries de la Societat Catalana de Terminologia; 9) [Publicació electrònica]

TRIAS, Mireia (cur.) (en premsa). *Terminologia per normalitzar la llengua*. Barcelona: Societat Catalana de Terminologia. (Memòries de la Societat Catalana de Terminologia; 10) [Publicació electrònica]



- * Col·lecció «Eines de Terminologia». Aplega els treballs guanyadors del Premi de la Societat Catalana de Terminologia:

GRESA BARBERO, Deva (2016). *Estudi d'implantació terminològica. La gastronomia del Pròxim Orient i el nord d'Àfrica als restaurants de Gràcia i l'Eixample de Barcelona*. Barcelona: Societat Catalana de Terminologia. (Eines de Terminologia; 1) [Publicació electrònica] [Premi de la Societat Catalana de Terminologia 2014]

MORTE SÁNCHEZ, Meritxell (2021). *La neologia de ciberseguretat en la llengua catalana*. Barcelona: Societat Catalana de Terminologia. (Eines de Terminologia; 2) [Publicació electrònica] [Premi ex aequo de la Societat Catalana de Terminologia 2020]

ARNAU GARCIA, Mariona (2021). *Adaptació del 'Diccionari casteller' del TERMCAT al públic infantil*. Barcelona: Societat Catalana de Terminologia. (Eines de Terminologia; 3) [Publicació electrònica] [Premi ex aequo de la Societat Catalana de Terminologia 2020]



Podeu consultar el web de la SCATERM per concretar i actualitzar dates i informacions. Pel que fa a l'adquisició de les nostres publicacions, podeu consultar el Portal de publicacions de l'IEC.

NORMES DE PRESENTACIÓ D'ORIGINALS

TERMINÀLIA és una revista d'accés obert, que no cobra als autors per la tramesa d'originals ni per la publicació. Els continguts de la revista estan subjectes a una llicència Reconeixement - No comercial - Sense obres derivades 3.0 de Creative Commons. Podeu enviar les vostres contribucions registrant-vos a la plataforma de la revista o per correu a terminalia@iec.cat.

La tramesa del manuscrit ha d'incloure una **carta de presentació** breu en què l'autor sol·licita l'avaluació del seu treball per a la possible publicació i indica la Secció a la qual l'adreça. La carta breu ha de contenir els elements següents: Títol del manuscrit en la llengua original i en anglès; declaració sobre l'adequació del treball presentat a la revista; nom complet de tots els autors amb la afiliació institucional, el correu electrònic i el número ORCID de cadascun; declaració sobre l'ordre de signatura dels autors de l'article (alfabètic o algun altre model). Opcionalment, s'hi pot afegir la referència del projecte o la institució que hi ha donat suport i el suggeriment de fins a quatre revisors adequats o bé descartats.

Extensió màxima. L'extensió màxima del document, incloent-hi imatges, resums i bibliografia, no pot excedir els 7.000 mots.

Llengües. Les contribucions han de ser preferentment en català o en anglès. En la secció Articles s'accepten també originals en espanyol i en francès.

Títol. El títol s'ha de presentar en la llengua original de l'article i en anglès, i ha de reflectir de manera clara, precisa i concisa la importància i el contingut del treball.

Resums. Tots els manuscrits adreçats a les seccions Articles i Dossier han d'incloure un resum estructurat, que exposi de manera clara i concisa els objectius, la metodologia, els resultats i les conclusions del treball. El resum s'ha de presentar en la llengua original de l'article; i també en català i en anglès, quan l'original no s'hagi redactat en cap d'aquestes llengües.

Paraules clau. Al final del resum l'autor ha de definir de cinc a deu paraules clau que ajudin a la indexació de l'article. Es poden utilitzar termes de tesaurus específics o generals elaborats per organismes internacionals de referència (com, per exemple, la UNESCO). Les paraules clau s'han de presentar en la llengua original de l'article i també en català i en anglès.

Notes. Les notes complementàries que s'incloguin en els originals han de ser notes al final de text numerades. Quan s'hagi de fer referència a un projecte, contracte o a un organisme que hagi finançat la recerca, cal que s'indiqui a la primera nota. Les notes també poden incloure agraïments o explicacions complementàries breus relacionades amb els continguts del text principal. Cal evitar que les notes continguin només referències bibliogràfiques, que s'han de citar dins del text i incloure a la llista de bibliografia final. Es recomana que no hi hagi més de 10 notes.

Gràfics. El material de suport gràfic (taules, gràfics, esquemes, fotografies, figures) inclòs en l'article s'ha de presentar també en arxius independents separats del text, preferentment en format TIFF o JPEG, i amb una resolució mínima de 300 ppp.

Bibliografia. Al final de l'escrit cal incloure l'apartat de bibliografia (també numerat), només amb les referències citades dins del text. Les referències bibliogràfiques que hagin de figurar dins el text han d'aparèixer entre parèntesis de la manera següent: (Cognom de l'autor/a, any, p. 000-000.). Les referències bibliogràfiques completes, ordenades alfabèticament, han de seguir els exemples següents: Cognom, Nom; Cognom, Nom(any). Títol del llibre. Número d'edició. Lloc d'edició: Editorial.

Cognom, Nom; Cognom, Nom(any). «Títol del capítol». A: Cognom, Nom; Cognom, Nom (any). Títol del llibre. Número d'edició. Lloc d'edició: Editorial, p. 000-000.

Cognom, Nom; Cognom, Nom(any). «Títol de l'article». Títol de la Revista, volum de la revista, número de la revista (tal com s'indica en l'original), p. 000-000.

Títol del diccionari. Número d'edició. Lloc d'edició: Editorial, any.

Cognom, Nom; Cognom, Nom(any). Títol [en línia]. Lloc d'edició: Editorial. [Consulta: dia mes any].

Sigles i abreviacions. Es recomana reduir al màxim l'ús de sigles en el cos de l'article. Quan n'hi hagi, cal assegurar que la primera vegada que hi surtin ho facin acompanyant la forma lèxica desenvolupada. Si s'usen sigles per referir-se a fonts bibliogràfiques, cal indicar-ho també en la bibliografia final. Per a la bibliografia, cal unificar les abreviatures amb les formes següents (tant si són en singular com en plural): ed. (editor/a o editors/ores), dir. (director/a o directors/ores), coord. (coordinador/a o coordinadors/ores), p. (pàgina o pàgines).

Format. Els articles s'han d'enviar en un format de processador de textos, preferentment MS Word o Open Office. L'arxiu ha d'incloure tot el contingut (títol, autoria, cos de l'article, imatges, taules, bibliografia, annexos, notes, resums i paraules clau) en la distribució prevista.

Tipus de lletra. Per al cos de l'article i la bibliografia, cal fer servir la lletra Times News Roman de 12 punts. Les citacions textuais, si són breus, van inserides en el text i es componen en regular rodona i entre cometes baixes («»); i, si són llargues, van entrades i es componen en regular rodona de la família Times News Roman d'11 punts. Les notes, si n'hi ha, van en Times News Roman de 10 punts.

Apartats. Els títols dels apartats han d'anar numerats (exemple: I, I.I, I.I.I, I.I.2...), amb el mateix cos de lletra, en negreta rodona. La primera línia de cada paràgraf ha d'anar entrada.

Interlineat. A un espai i mig, en tots els casos. Els paràgrafs no se separen amb cap interlineat extra. Cal deixar una línia en blanc entre títol i text i dues línies entre text i títol.



Crida d'articles

Els articles, que han de ser originals i inèdits, són avaluats per revisors externs designats pel Consell de Redacció de la revista. En cas que s'accepti la proposta, els autors poden rebre suggeriments de modificació.

Les propostes han de seguir les normes de presentació d'originals de la revista.

Els autors han d'enviar les seves propostes per correu electrònic a terminalia@iec.cat. Alternativament, es poden registrar en el sistema electrònic de gestió de la revista (Open Journal System), a <http://revistes.iec.cat> o a <http://terminalia.iec.cat>, i penjar-hi les trameses.

Per al número 30 (desembre 2024)

Tramesa d'articles: fins al 31 de desembre de 2023

Resultats de la selecció: fins 28 de febrer de 2024

Versió final: fins al 31 de març de 2024

Publicació de la revista: desembre de 2024

Per al número 31 (juny 2025)

Tramesa d'articles: fins al 30 de juny de 2024

Resultats de la selecció: fins al 15 de setembre de juliol

Versió final: fins al 30 de setembre de 2024

Publicació de la revista: juny de 2025



Protecció de dades personals

L'Institut d'Estudis Catalans (IEC) compleix el que estableix el Reglament general de protecció de dades de la Unió Europea (Reglament 2016/679, del 27 d'abril de 2016). De conformitat amb aquesta norma, s'informa que, amb l'acceptació de les normes de publicació, els autors autoritzen que les seves dades personals (nom i cognoms, dades de contacte i dades de filiació) puguin ser publicades en el corresponent volum de *Terminàlia*.

Aquestes dades seran incorporades a un tractament que és responsabilitat de l'IEC amb la finalitat de gestionar aquesta publicació. Únicament s'utilitzaran les dades dels autors per a gestionar la publicació de *Terminàlia* i no seran cedides a tercers, ni es produiran transferències a tercers països o organitzacions internacionals. Un cop publicades a *Terminàlia*, aquestes dades es conservaran com a part del registre històric d'autors. Els autors poden exercir els drets d'accés, rectificació, supressió, oposició, limitació en el tractament i portabilitat, adreçant-se per escrit a l'Institut d'Estudis Catalans (carrer del Carme, 47, 08001 Barcelona), o bé enviant un correu electrònic a l'adreça dades.personals@iec.cat, en què s'especifiqui de quina publicació es tracta.

