

Garbell: l'avaluador automàtic de neologismes catalans¹

JUDIT FREIXA AYMERICH

Universitat Pompeu Fabra

ORCID: 0000-0002-1440-5514

judit.freixa@upf.edu

Judit Freixa Aymerich és doctora

en Filologia (Universitat de Barcelona, 2002) i professora titular del Departament de Traducció i Ciències del Llenguatge de la Universitat Pompeu Fabra (UPF), on imparteix les matèries de lèxic, terminologia i neologia. És directora de l'Observatori de Neologia (UPF) i investigadora principal de diversos projectes competitiu sobre neologia i lexicografia: GARBELL: Analitzador de la diccionariabilitat dels neologismes catalans (IEC: PRO2018-S04-FREIXA); NADIC: Neologismes per a l'actualització del diccionari (PR2015-S04-FREIXA); LEXICAL: Neología y diccionario. Análisis para la actualización lexicográfica del español (PID2020-118954RB-I00), i NEÓMETRO: Medición de la neologicidad y la diccionariabilidad (FFI2016-79129-P). És membre del grup IULATERM i la majoria de les seves publicacions ha aparegut en revistes o editorials d'impacte (*Treballs de Sociolingüística Catalana*, *Terminàlia*, *Sendebar*, *Caplletra*, *Revista de Llengua i Dret*, *Revista Española de Lingüística*, *Terminology*, *Meta*, *Signos*).



Resum

El Garbell és una eina que avalua neologismes i determina si són més o menys diccionaritzables tenint en compte criteris d'ús, documentals i lingüístics. El resultat, en format web, públic i gratuït, és una eina innovadora que serveix de diccionari complementari i progressiu per al català.

PARAULES CLAU: neologisme; diccionari; diccionariabilitat; català; criteris

Abstract

Garbell: the automatic catalan neologisms analyzer

Garbell is a tool that evaluates neologisms and determines their degree of interest for lexicographic purposes, taking into account usage, documentary and linguistic criteria. The result, in web format, open access and free, is an innovative tool that serves as a complementary and progressive dictionary for Catalan

KEYWORDS: neologism; dictionary; diccionariabilitat; Catalan; criteria

TERMINÀLIA 26 (2022): 7-16 · DOI: 10.2436/20.2503.01.179

Data de recepció: 02/10/2022. Data d'acceptació: 10/11/2022

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

1 Introducció i objectius

Els parlants de català utilitzem diàriament paraules que no apareixen al diccionari normatiu, el *Diccionari de la llengua catalana* de l'IEC (DIEC, en endavant). Algunes d'aquestes paraules són freqüents, s'utilitzen en tots els registres de la llengua i no hi ha cap motiu per considerar-les incorrectes. Són paraules diccionaritzables, paraules que podrien aparèixer al diccionari. En canvi, n'hi ha d'altres que encara no han demostrat si són realment necessàries per als parlants i haurà de passar un temps abans no es puguin diccionaritzar o rebutjar; són paraules que podem considerar prediccionaritzables. Finalment, els parlants també fem servir paraules que no són al diccionari ni hi han de ser. Són paraules no diccionaritzables que sovint utilitzem per motius estilístics.

El Garbell és una eina que avalua aquestes paraules que no són al diccionari i determina si són més o menys diccionaritzables; ho fa tenint en compte criteris d'ús, documentals i lingüístics. La majoria d'aquests criteris s'apliquen automàticament, i per això diem que el Garbell és un avaluador automàtic de la diccionariabilitat dels neologismes catalans, però encara hi ha algun criteri que s'ha d'aplicar manualment i, per això, no podem considerar-la una eina totalment automàtica.

El Garbell consta d'un programa informàtic que avalua els neologismes i una base de dades pública on es pot consultar el resultat general de la diccionariabilitat de cada neologisme i el resultat específic de cada criteri. Ha estat desenvolupat per investigadores de l'Observatori de Neologia del grup IULATERM (Universitat Pompeu Fabra) i ha rebut finançament de l'Institut d'Estudis Catalans (IEC) entre els anys 2018 i 2021, i de la Universitat Pompeu Fabra.

En aquest article presentem el funcionament i els resultats del Garbell, però primer volem emmarcar-lo en la recerca actual sobre la relació entre neologismes i diccionaris i, després, valorar les limitacions actuals d'aquesta línia de recerca i les possibilitats per al futur.

2 Diccionariabilitat i criteris de diccionarització

És ben sabut que quan es recullen neologismes lexicogràfics (és a dir, paraules que es documenten en l'ús i que no estan recollides als diccionaris), no tots tenen les mateixes possibilitats de ser incorporats al diccionari en un futur. De fet, molts no són ni tan sols neologismes pròpiament dits, perquè també trobem paraules ja existents però especialitzades, o transparents, o dialectals, etc., o ocasionalismes, aquelles unitats d'aparició única i de funció estilística. I dins dels neologismes pròpiament dits, els que són realment mots nous, n'hi ha de més diccionaritzables que d'altres.

Diccionaritzar una paraula és incorporar-la al diccionari. Normalment, la diccionarització es refereix a

la incorporació en un diccionari monolingüe de referència, però podria no ser així. Perquè una paraula s'incorpori al diccionari ha de ser diccionaritzable, és a dir, ha de complir els criteris d'inclusió fixats per l'empresa o l'organisme que s'ocupa d'actualitzar el diccionari i, a més, no ha de complir cap dels criteris d'exclusió i que la fan rebutjable. En aquest sentit, cal remarcar que les diferents tradicions lexicogràfiques tenen també diferents posicions davant la incorporació de neologismes; de manera genèrica, podem afirmar que la lexicografia romànica és més reguladora que l'anglòfona, per exemple. En qualsevol cas, les paraules són més o menys diccionaritzables segons els criteris de diccionarització que es compleixin o s'incompleixin, i la diccionariabilitat és la qualitat de ser diccionaritzable.

Amb l'estudi de la diccionariabilitat es busca la manera d'objectivitzar l'interès lexicogràfic que pot tenir una paraula per a un diccionari, havent definit el tipus de diccionari que es vol actualitzar, ja que poden tenir ben poc en comú l'actualització d'un diccionari de neologismes i d'un diccionari prescriptiu. La reflexió sobre els criteris per a la diccionarització dels neologismes ha format part de la pràctica lexicogràfica des dels seus orígens, però des de fa uns anys existeix també com una línia de recerca teòrica en si mateixa, iniciada en l'àmbit anglòfon per Barnhart (1985) i continuada amb els treballs d'Algeo (1993), Sheidlower (1995), Ishikawa (2006) i O'Donovan, Ruth; O'Neill, Mary (2008), entre altres. Aquests treballs centraven l'atenció en els criteris d'ús, i molt específicament la freqüència d'ús, per determinar si una paraula és diccionaritzable o no. La freqüència s'hi estudia com un criteri complex i, més enllà del nombre total d'ocurrències d'un mot en un corpus, es té en compte la diversitat textual, geolectal i de registre. En l'àmbit hispanòfon, els treballs d'Adelstein i Freixa (2013) i de Sánchez Manzanares (2013) van mostrar l'interès per aquesta línia, a la qual es va sumar molt aviat el català amb treballs derivats del projecte NADIC (vegeu l'apartat 3). En l'estudi de la diccionariabilitat cal tenir també en compte treballs importants com el de Nam et al. (2016) per al coreà, Bernal et al. (2020 i 2022) per al castellà, i Klosa i Wolfer (2020) per a l'alemany.

3 Antecedents del Garbell: el projecte NADIC

El projecte Garbell beu directament del [NADIC](#)² ('Neologismes per a l'actualització del diccionari normatiu'), un projecte anterior també liderat per l'Observatori de Neologia en què van participar investigadors dels observatoris de neologia de NEOXOC,³ la xarxa que aplega els observatoris de neologia dels territoris de parla catalana i que va rebre finançament de l'IEC durant els anys 2015-2017. L'objectiu del NADIC era la selecció dels mil neologismes més diccionaritzables a partir d'una selecció inicial de 30.000 neologis-

mes detectats pels observatoris de neologia catalans i disponibles a la base de dades pública BOBNEO.⁴ Per poder seleccionar les unitats de manera objectiva, es va elaborar la bateria de criteris que es pot veure a la taula 1, basada en les aportacions de la bibliografia i l'estudi de les dades.

Els criteris del NADIC es presenten agrupats en tres blocs: criteris d'ús, criteris lingüístics i criteris de documentació. Al seu torn, els criteris poden ser d'exclusió, de manera que la unitat que no compleix el criteri es descarta automàticament, o d'inclusió, de manera que la unitat que compleix el criteri suma aquell aspecte dins la matriu de diccionarització. Són criteris d'exclusió els set criteris ombrejats a la taula i criteris d'inclusió els quinze no ombrejats.

El Garbell s'ha beneficiat de l'experiència adquirida en el NADIC, tant pel que fa a la identificació dels criteris com a la informació derivada de la seva aplicació. Per exemple, amb el NADIC es va observar que alguns criteris tenen més potencial discriminant que d'altres, perquè afecten un major nombre d'unitats. Així, els criteris amb major impacte són, per aquest ordre: la freqüència alta, la presència en diccionaris d'altres llengües, en diccionaris especialitzats i en altres diccionaris catalans. El projecte NADIC també va permetre constatar que, un cop seleccionat el miler de neologismes més diccionaritzables, una part important dels neologismes que quedaven fora de la selecció eren també altament diccionaritzables (com ara *agrocombustible*, *ninja*, *multiètnic -a*, *neoconservador -a* o *reaprofitar*), de manera que es va veure la necessitat de no tancar la línia d'investigació amb el NADIC i de pensar en una eina més sostenible, que funcionés amb menys intervenció humana i que anés actualitzant els resultats per poder donar compte de la dinamicitat de la llengua.

La majoria de criteris del NADIC s'han pogut traslladar al Garbell, és a dir, s'han pogut automatitzar.

Altres criteris, però, s'han hagut de desestimar o s'han d'aplicar manualment, com veurem en l'apartat següent, on també ens referirem a altres criteris identificats per al Garbell.

4 El Garbell

El Garbell no detecta neologismes, sinó que analitza els neologismes catalans ja detectats i registrats al BOBNEO, la base de dades dels observatoris de neologia, i n'avalua la diccionariabilitat. Els neologismes registrats a BOBNEO segueixen el criteri lexicogràfic i, doncs, són paraules documentades en l'ús (especialment en premsa, però també en revistes, Twitter i textos radiofònics) i no recollides al DIEC. Amb aquest criteri es recullen els neologismes pròpiament dits (unitats recents en l'ús com *violència sicària*, *sindèmia*, *preprint* o *postmemòria*), però també altres unitats que no són pròpiament neològiques i que no són al diccionari per altres criteris que no són la novetat: són especialitzades, col·loquials, dialectals o transparents.

El Garbell no valora la necessitat social d'aquests neologismes: si són a l'ús, és que són necessaris d'alguna manera. Però sí que té en compte aquesta necessitat (mesurada amb la freqüència) i altres criteris lingüístics i documentals, per determinar si són candidats més o menys bons per al diccionari.

4.1 Criteris i algorisme

Com en el NADIC, els criteris que s'apliquen per valorar la diccionariabilitat es reparteixen en tres blocs: criteris d'ús, lingüístics i documentals. En la taula 2 es pot veure la distribució dels criteris, que s'expliquen més avall (com abans, els criteris ombrejats són criteris d'exclusió).

Criteris d'ús	Criteris lingüístics i temàtics	Criteris de documentació
I. actualitat	II. formació	18. presència en dic. especialitzats
2. presència	12. variant	19. presència en dic. generals prioritzats
3. estabilitat	13. compleció de sèrie derivativa	20. presència en dic. generals no prioritzats
4. extensió	14. compleció de sèrie semàntica	21. presència en dic. d'altres llengües
5. localisme	15. compleció de camp temàtic	22. presència en corpus
6. freqüència elevada	16. impacte	
7. freqüència molt elevada	17. impredictibilitat	
8. representativitat		
9. orientació		
10. validació		

TAULA 1. Criteris de diccionarització proposats en el projecte NADIC

Criteris d'ús (9)	Criteris lingüístics i temàtics (4)	Criteris de documentació (11 fonts)
actualitat	construcció regular	NEOLOTECA
presència	(no) variant	CERCATERM
estabilitat	impacte	GDLC
extensió d'ús	(no) possible error	D62*
(no) localisme		VOX
freqüència alta		DNV
freqüència molt alta		Silencis del DIEC*
representativitat geolectal		Ésadir*
ús (no) marcat		DEL*, ROB*, ZING*

TAULA 2. Criteris de diccionarització del Garbell

4.1.1 Criteris d'ús

Amb els criteris d'ús s'aconsegueix que el Garbell prioritzi els candidats actuals més freqüents i representatius.

Amb el primer criteri es prioritza l'actualitat dels candidats, perquè es rebutgen els que no han aparegut en l'ús durant els darrers cinc anys. Així, neologismes que van ser molt freqüents durant una època però que han perdut la vigència, ja no segueixen el procés per aquest criteri d'exclusió que assegura l'essència del neologisme, que és l'actualitat. Aquest criteri es combina amb el tercer, l'estabilitat, un criteri d'inclusió per prioritzar els candidats que ja han pogut demostrar que són d'interès per als parlants perquè tenen un mínim de dos anys d'antiguitat. Amb aquest criteri es pretén no precipitar la diccionarització de neologismes molt joves encara i que podrien deixar d'estar presents en l'ús en poc temps. Així, amb la combinació d'aquests dos criteris es prioritzen els candidats recents, però ja mínimament assentats en l'ús.

Tres dels criteris d'aquest bloc (segon, sisè i setè) es refereixen a la freqüència, el criteri sobre el qual hi ha més consens a la bibliografia: amb el criteri de presència s'exclouen els candidats que no arriben a un mínim de quatre ocurrences en els últims cinc

anys, per excloure automàticament els candidats que no presenten indicis d'estabilització en l'ús, i amb els altres dos, relatius a la freqüència elevada i molt elevada, es prioritzen els candidats més implantats en l'ús.

L'extensió de l'ús es valora amb tres criteris més: amb l'extensió textual es pretén prioritzar la selecció de candidats que es documenten en diferents tipus de textos), però de moment és un criteri per excloure els neologismes estrictament orals, pel seu caràcter poc diccionaritzable (per exemple, en l'oralitat es produeix una gran quantitat d'interferències amb el castellà). D'altra banda, amb el cinquè criteri es pretén excloure els candidats que són modes, localismes o paraules testimonis d'una època perquè els diccionaris prefereixen no donar entrada a aquest tipus de mots que els parlants deixen d'utilitzar al cap d'un temps. El Garbell en pot detectar una part, concretament els que es poden parafrasejar com a «relatiu a x» on x és el nom propi d'una institució, una persona o un lloc, perquè venen marcades amb una nota des de l'anàlisi que es realitza en la cadena de treball de l'Observatori de Neologia.

Finalment, amb el criteri de representativitat geolectal es marquen les unitats que són representatives a una part del territori lingüístic i que difícilment aconseguiran una gran freqüència en l'ús perquè els tex-

Criteris d'ús (9)	
actualitat	Documentat durant els últims 5 anys.
presència	Té almenys 4 ocurrences.
estabilitat	La primera ocurrencia té almenys 2 anys d'antiguitat.
extensió d'ús	No és exclusivament oral.
(no) localisme	No té nota gRc.
freqüència alta	Té un mínim de 10 ocurrences
freqüència molt alta	Té un mínim de 25 ocurrences
representativitat geolectal	Només es documenta en una font que no és de BCN.
ús (no) marcat	No apareix amb marques tipogràfiques

TAULA 3. Criteris d'ús del Garbell

tos que es buiden no els solen recollir; i amb l'últim criteri, ús (no) marcat, es pretén tornar a focalitzar en les unitats més estabilitzades en l'ús, les que ja apareixen sense marques tipogràfiques de suport, un criteri implantat en el Garbell (no prové del NADIC) pel seu caràcter automàtic.

A més d'aquests nou criteris, en la fase de validació manual de les dades se n'apliquen d'altres, com el criteri d'orientació de l'ús, que considera més diccionaritzables els neologismes amb una forma gràfica que pot generar dubtes, com ara la duplicació de la lletra r, la e epentètica, la l·l, etc.

4.1.2 Criteris lingüístics

Els criteris lingüístics són els més difícilment automatitzables i per això és actualment la part menys desenvolupada del Garbell.

Per prioritzar els neologismes més predictibles i transparents, el Garbell té en compte, en primer lloc, si es tracta d'un mot construït (derivats, compostos, sintàctics i abreviacions) i se li suposa la regularitat, perquè els neologismes amb una formació transgressora ja han estat exclosos pel criteri de freqüència mínima que hem anomenat *presència* en els criteris d'ús. D'altra banda, també s'avalua l'impacte sobre el diccionari, ja que l'addició d'una nova accepció o una subentrada és menys costosa en lexicografia que una entrada completa nova, de manera que els neologismes semàntics i sintagmàtics es consideren més diccionaritzables. Amb la suma d'aquests dos criteris, queden penalitzats els manlleus i els neologismes formats per processos menys productius com l'acronímia i diferents tipus de canvis sintàctics. Com es pot veure, està ombrejat el criteri de variant, amb què s'exclouen tots els neologismes ja recollits al diccionari amb una altra forma (variants gràfiques i ortogràfiques, però no lèxiques).

Els criteris addicionals que es tenen en compte en la validació manual són els criteris no automatitzables que es van aplicar en el projecte NADIC (compleció de sèrie derivativa, semàntica, de camp temàtic i impredecibilitat semàntica), però també altres criteris particulars que es van aïllant i que poden ser automatitzats, com ara la no diccionariabilitat de tots els càrrecs i oficis amb el prefix *ex-*.

4.1.3 Criteris documentals

Quan un diccionari incorpora una paraula a la nomenclatura dona fe de la necessitat que tenen d'aquella paraula els parlants d'una llengua. Per això, el Garbell té en compte el criteri de la presència d'un neologisme en altres diccionaris. En primer lloc, es consulten obres especialitzades (Neoloteca i Cercaterm) i, després, altres diccionaris generals del català que, pel fet de no ser normatius, tenen un caràcter més obert i descriptiu: *Gran diccionari de la llengua catalana* (GDLC), *Gran diccionari 62 de la llengua catalana* (GD62), *Diccionari manual de la llengua catalana* (VOX). També es consulta el *Diccionari normatiu valencià* i altres fonts complementàries per al català com l'*ésAdir* (el portal lingüístic de la Corporació Catalana de Mitjans Audiovisuals) i el blog de Salvanyà *Silencis del DIEC*. Finalment, també es té en compte el tractament que han rebut els neologismes en diccionaris generals de referència d'altres llengües romàniques: *Diccionario de la lengua española* (DLE), *Lo Zingarelli* (ZING) i *Le Grand Robert* (ROB).

En total, doncs, les fonts d'aquest bloc de criteris són onze i la majoria es consulta automàticament (si bé els neologismes semàntics i sintàctics queden sempre fora de les consultes automàtiques perquè el sistema només pot llegir les formes i no les definicions). Els diccionaris de les altres llengües i les fonts que no disposen d'un lemarí automatitzat s'han de consultar manualment en la fase de validació i, mentrestant, el programa atribueix la categoria de *pendent* als neologismes per als quals no té prou informació documental per determinar-ne l'estat de diccionariabilitat.

4.1.4 Algoritme de diccionariabilitat

Els criteris que acabem de veure no se sumen linealment per atribuir un valor final de diccionariabilitat, sinó que aquest valor s'obté per l'aplicació d'un algoritme que s'aplica en dos nivells. En el primer nivell s'obté un resultat per a cada bloc de criteris. Per exemple, en els criteris d'ús: per obtenir un valor alt, el neologisme ha de complir tres o més criteris d'inclusió (dels cinc existents); obté un valor mitjà si en compleix un o dos, i un valor baix si no en compleix cap.

En el segon nivell, l'algoritme combina el resultat obtingut en cada bloc de criteris per atorgar el valor

Criteris lingüístics (4)	
construcció regular	És un mot construït (PREF, SUF, CULT, COMP, LEX, CONV, ABR).
(no) variant	No té la nota 2a.
impacte	És un neologisme semàntic o sintagmàtic.
(no) possible error	No té la nota 3.

TAULA 4. Criteris lingüístics del Garbell

final, que podrà ser: diccionaritzable, prediccionaritzable, no diccionaritzable o pendent. La combinació de criteris no és lineal, però, en síntesi, són diccionaritzables les unitats que almenys en dos dels blocs tenen un resultat positiu i són prediccionaritzables les unitats que almenys en dos dels blocs tenen un resultat mitjà. Obtenen el resultat de no diccionaritzables els neologismes que compleixen algun criteri d'exclusió i, finalment, el sistema deixa com a pendents els neologismes que no tenen un resultat definitiu en el bloc de criteris documentals, per falta d'informació.

4.2 Resultats

El Garbell és una eina pública consultable a <http://garbell.upf.edu/> des del setembre de 2022. Per tal d'afinar els resultats, durant el primer any s'estan sotmetent a revisió manual tots els resultats i s'estan completant les fonts que no es poden consultar automàticament. Això permetrà revisar els criteris i l'algorisme i aconseguir que en una versió futura, el programa sigui més autònom i fiable. S'espera que al final del procés, el Garbell ofereixi més de 10.000 neologismes avaluats i classificats segons el seu estat de diccionariabilitat.

El Garbell es pot consultar mitjançant les cerques bàsiques de les bases de dades. Per exemple, es poden buscar tots els neologismes que comencin per una lletra o segment. En la figura 1 es pot veure l'exemple de la cerca de neologismes ja publicats que comencen per la lletra s.

Els neologismes destacats amb el fons verd són diccionaritzables i els altres són no diccionaritzables o prediccionaritzables, segons s'indiqui. Com es pot veure, les variants gràfiques i ortogràfiques apareixen unides i tenen un valor de diccionariabilitat independent. Així, la forma *sancta sanctorum* (sense accent) és diccionaritzable i, en canvi, la forma amb accent, no ho és. De fet, la forma diccionaritzable és més freqüent i, a més, ja està recollida al Cercaterm, al GDLC, GD62, al DNV i al ZING (i l'altra forma no està recollida enlloc). De vegades, els grups de variants són molt nombrosos i, de vegades, no n'hi ha cap de diccionaritzable.

A la figura 1 hi ha neologismes semàntics, unitats com *sacsejar* o *segell*, que tenen un significat diferent del que apareix ja recollit en el diccionari. També hi ha unitats derivades, com *sancionador -a* i *salafista*, diccionaritzables, totes amb freqüència molt alta i ja recollides en cinc fonts documentals, o *seguidista* i *sagarrià -ana* (prediccionaritzables); abreviacions lèxiques, com *segurata*, que el Garbell considera prediccionaritzable.

Neologismes que comencin per "s"

S'han trobat 286 resultats amb un total de 333 neologismes.

sabatòfon <i>m</i>	No diccionaritzable	sacsejar <i>v tr</i>	Diccionaritzable
sagarrià -ana <i>adj</i>	Prediccionaritzable	sal grossa <i>f</i>	Prediccionaritzable
salafista <i>adj</i>	Diccionaritzable	salero <i>m</i>	No diccionaritzable
salut mental <i>f</i>	Diccionaritzable	sancionador -a <i>adj</i>	Diccionaritzable
sanfermines <i>m pl</i>	No diccionaritzable	sashimi <i>m</i>	Diccionaritzable
sauvignon <i>m</i>	Prediccionaritzable	script <i>m</i>	Diccionaritzable
sectorització <i>f</i>	Diccionaritzable	secundari secundària <i>m i f</i>	Prediccionaritzable
segell <i>m</i>	Diccionaritzable	segellar <i>v tr</i>	Diccionaritzable
seguidisme <i>m</i>	Prediccionaritzable	seguidista <i>adj</i>	Prediccionaritzable
segurata <i>m i f</i>	Prediccionaritzable	seitan <i>m</i>	Diccionaritzable

FIGURA 1. Cerca per segment inicial (lletra s)

interanual *adj*

Diccionaritzable			Any de la primera ocurrència: 1989									
<table border="1"> <tr> <td>Criteris d'ús</td> <td>Criteris lingüístics</td> <td>Criteris documentals</td> </tr> <tr> <td> Alt</td> <td> Alt</td> <td> Mitjà</td> </tr> <tr> <td>VEURE</td> <td>VEURE</td> <td>VEURE</td> </tr> </table>	Criteris d'ús	Criteris lingüístics	Criteris documentals	Alt	Alt	Mitjà	VEURE	VEURE	VEURE	Nombre total d'ocurrències: 340		
Criteris d'ús	Criteris lingüístics	Criteris documentals										
Alt	Alt	Mitjà										
VEURE	VEURE	VEURE										

CONTEXT

Pel que fa al conjunt de les pensions -jubilatció, viudetat, orfanat, incapacitat i favors a familiars-, se n'han concedit un total d'1,74 milions, un 0,48% més en termes *interanuals*.
[El Diari de Girona, 29/08/2021]

[Mostrar 339 contextos addicionals.](#)

OCURRÈNCIES PER ANY

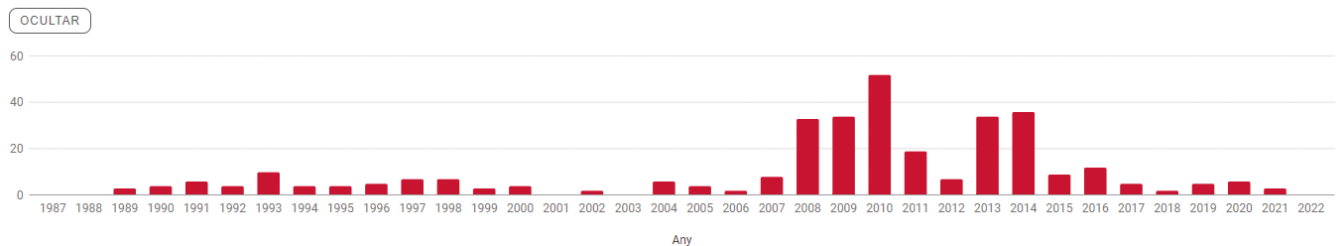


FIGURA 2. Cerca per mot (interanual)

zable perquè, tot i la freqüència alta, només apareix al GD62 i a SIL (pel fet de ser una interferència present especialment en l'ús oral) i també abreviacions sintagmàtiques, com *secundària* (d'*escola secundària*). També hi ha unitats sintagmàtiques com *salut mental* i *sal grossa* i manlleus com *sashimi*, *script*, *salero* o *sauvignon* (no tots diccionaritzables).

Amb l'adjectiu *interanual* mostrem, en la figura 2, la informació que dona el Garbell per a les consultes per mot. A banda de la capçalera amb el mot i la categoria gramatical, el programa mostra una síntesi en una pastilla on s'indica l'estat de diccionariabilitat (diccionaritzable, en aquest cas) i el valor de cada bloc de criteris. Al costat s'ofereix la informació de l'any de la primera ocurrència i la informació del nombre total

d'ocurrències i a sota, un context d'ús (amb font i data) i un botó per consultar altres contextos. Al final de tot, hi ha l'histograma on es mostra l'evolució de l'ús al llarg dels anys.

Si cliquem l'opció 'veure' en la pastilla inicial, es despleguen els resultats obtinguts en cada criteri de cada bloc. En aquest cas, es compleixen set dels nou criteris d'ús, tres dels quatre criteris lingüístics (per això en tots dos casos el valor és alt en la pastilla de síntesi), però el valor de la part documental no és alt perquè només apareix a tres fonts.⁵

Observant els resultats automàtics, podem concloure que el Garbell distingeix clarament els diferents estats de diccionariabilitat en la majoria de casos. Són diccionaritzables els neologismes equiparables a les

CRITERIS D'ÚS

Actualitat		Localisme		Freqüència molt alta	
Presència		Estabilitat		Representativitat	
Extensió		Freqüència alta		Ús no marcat	

CRITERIS LINGÜÍSTICS

Variet		Formació	
Impacte	No aplicable	Error	

CRITERIS DOCUMENTALS

Neoloteca		VOX		ZING	
Cercaterm		DNV		ROB	
GDLC		Silencis		ALT	
GD62		DLE			

FIGURA 3. Criteris desplegats (interanual)

paraules ja diccionaritzades i que compleixen tots o la majoria dels requisits per ser-ho, que ja tenen un recorregut en l'ús social i que no mostren cap característica que pugui qüestionar-ne la sanció lexicogràfica. En canvi, són prediccionaritzables els neologismes que encara no tenen un recorregut suficient per ser diccionaritzables i, alhora, no presenten cap característica que els impedeixi ser-ho en un futur proper. Finalment, són no diccionaritzables els neologismes que es troben a l'inici del recorregut (i, per tant, no compleixen el mínim de requisits) o que ja han avançat en aquest recorregut, i ja han mostrat característiques que els impedeixen ser diccionaritzables.

Ara bé, el Garbell també ofereix alguns resultats sorprenents i que han fet necessària la revisió manual de la totalitat de les dades. A la figura 4 podem veure el cas de *secallona*, un mot clarament no neològic però que el Garbell ha analitzat perquè és un neologisme lexicogràfic (és a dir, apareix en l'ús i no està recollit en el DIEC). Com es pot veure, els criteris d'ús donen un resultat positiu, i també els criteris lingüístics, però el criteri documental converteix el mot en prediccionaritzable, per tal com només apareix recollit en una font.

Semblantment, és sorprenent el resultat de baixa diccionariabilitat del verb *reaprofitar*, estable en l'ús des dels anys 90, amb una freqüència d'ús molt elevada i sense cap limitació lingüística, però que no apareix en cap font (segurament pel seu caràcter transparent). Aquests i altres exemples amb resultat sorprenent indiquen que el criteri documental, que sol funcionar adequadament en la discriminació de la diccionariabilitat dels neologismes, resulta excessiu en un petit

percentatge de casos, quan moltes obres han coincidit en un oblit lexicogràfic.

4.3 Valoració del Garbell

La utilitat del Garbell és inqüestionable, tot i no ser encara una eina que pugui separar del tot automàticament el gra (i diferents qualitats de gra) de la palla. Fins a l'actualitat, no coneixem cap eina, en cap llengua, que pugui avaluar, com el Garbell, la diccionariabilitat de neologismes de manera automàtica i molt menys fer-ho amb una bateria de criteris on es combinen diferents angles d'anàlisi.

La principal limitació del Garbell és, actualment, la necessitat de validació manual de les dades, que hauria de desaparèixer (o, almenys, minimitzar-se) de cara al futur. De moment, el programa no fa excepcions i exclou unitats totalment diccionaritzables segons el sentit comú i el coneixement especialitzat, com hem il·lustrat abans; aquesta limitació es podrà resoldre amb la cerca a corpus i amb un etiquetatge més afinat de les dades; alguns dels errors del Garbell tenen a veure amb incoherències del BOBNEO que dispersen els resultats, i aquest és un dels aspectes que s'està resolent amb la revisió de la base de dades. D'altra banda, alguns blocs de neologismes, com els semàntics, s'escapen encara de qualsevol tractament automàtic fiable.

De fet, precisament perquè ens trobem en un terreny desconegut, és necessària la revisió manual de les dades. Només així podem observar quins aspectes del funcionament de Garbell són sòlids i quins cal replan-tejar o desenvolupar encara més.

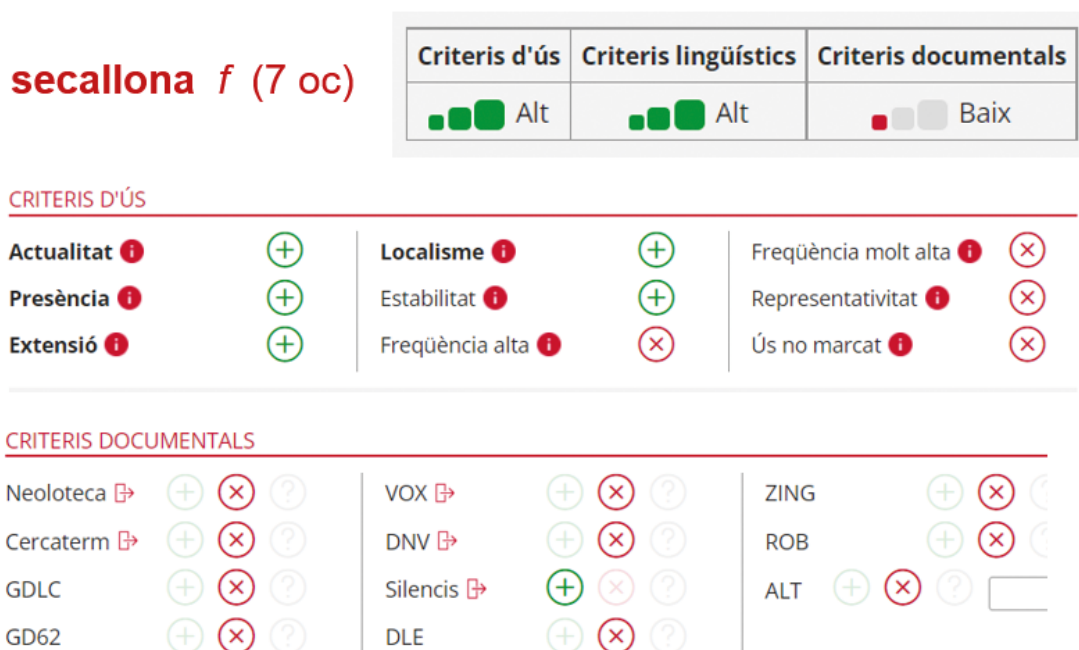


FIGURA 4. El pes del criteri documental (*secallona*)

5 Conclusions

El resultat del Garbell és un recurs lingüístic que serveix com a diccionari complementari, progressiu i que orienta l'ús. És complementari perquè recull precisament allò que no inclou el diccionari de referència; és progressiu perquè va més enllà d'una concepció dicotòmica i situa els candidats a neologismes en diferents punts més o menys allunyats de la meta que és incorporar-los al diccionari, i orienta l'ús, perquè informa també de les unitats excloses.

El Garbell està totalment desenvolupat informàticament i està obert al públic des del setembre de 2022, però encara està en fase de validació i anirà incrementant el nombre de neologismes publicats fins a arribar a un nombre superior als 10.000. Quan s'hagi tancat l'etapa de validació manual, la revisió del Garbell (dels criteris i de l'algoritme) permetrà un funcionament més autònom i anirà incrementant anualment les dades actualitzades sobre la diccionariabilitat dels neologismes catalans. ✨

Bibliografia

- [BOBNEO] Banc de dades dels observatoris de neologia (Barcelona: Institut de Lingüística Aplicada, Universitat Pompeu Fabra, 1989-), <<http://obneo.iula.upf.edu/bobneo/index.php>>, consultat: setembre del 2022.
- [CERCATERM] Cercaterm (Barcelona: Termcat, Centre de Terminologia), <<http://www.termcat.cat/es/Cercaterm/>>, setembre del 2022.
- [GD62] López del Castillo (dir.) *Gran diccionari 62 de la llengua catalana*. (Barcelona: Edicions 62, 2000).
- [DIEC] Institut d'Estudis Catalans, *Diccionari de la llengua catalana* (Barcelona: Edicions 62; Enciclopèdia Catalana, 2007).
- [DLE] Real Academia Española, *Diccionario de la lengua española*, 23rd ed. (Madrid: Real Academia Española, 2014), <<http://www.rae.es/rae.html>>, setembre del 2022.
- [DNV] Acadèmia Valenciana de la Llengua, *Diccionari normatiu valencià* (València: Acadèmia Valenciana de la Llengua, 2014), <<https://www.avl.gva.es/lexicval/>>, setembre del 2022.
- [GDLC] *Gran diccionari de la llengua catalana* (Barcelona: Enciclopèdia Catalana, 1998), <<http://www.diccionari.cat/>>, setembre del 2022.
- [LGR] *Le Grand Robert de la langue française*, 4.1 version (Paris: Dictionnaires Le Robert, 2017), <<https://gr.bvdep.com>>, setembre del 2022.
- [LZ] Lo Zingarelli 2018. *Vocabolario della lingua italiana* (Bologna: Zanichelli, 2017).
- [NEOLOTECA] Neoloteca (Barcelona: Termcat, Centre de Terminologia), <<http://www.termcat.cat/es/Neoloteca/>>, setembre del 2022.
- [VOX] *Diccionari manual de la llengua catalana VOX* (Barcelona: Larousse Editorial, 2015).
- ADELSTEIN, Andreína; FREIXA, Judit (2013). «Criterios para la actualización lexicográfica a partir de datos de observatorios de neología». Presentació inèdita, Congreso Internacional El Diccionario: neología, lenguaje de especialidad, computación, Ciudad de México (Mexico), 28-30th October 2013. Consultable a: <https://repositori.upf.edu/handle/10230/34891>
- ALGEO, John. (1993). «Desuetude among new English words». *International Journal of Lexicography* 6 (4). p. 281-293.
- BARNHART, David K. (1985). «Prizes and pitfalls of computerized searching for new words for dictionaries». *Dictionaries* 7. p. 253-260.
- BERNAL, Elisenda; FREIXA, Judit; TORNER, Sergi. (2020). «Criterios para la diccionarización de neologismos: de la teoría a la práctica». *Signos* 53 (104). p. 592-618.
- BERNAL, Elisenda; FREIXA, Judit; TORNER, Sergi (2022). *La neología del español: del uso al diccionario*. Iberoamericana.
- FREIXA, Judit; TORNER, Sergi (2020). «On the Dictionarization of New Words in Spanish». *Dictionaries* 41 (1). p. 131-153.
- FREIXA, Judit; MARTINES, Josep; GUARDIOLA, M. Isabel; Montané, M. Amor (2022). *The Dictionarisation of Catalan Neologisms*. Peter Lang.
- ISHIKAWA, Shin'ichiro (2006). «When a word enters the dictionary: A data-based analysis of neologism». A: JACET Society of English Lexicography (ed.). *English Lexicography in Japan*. Bunkyo-ku: Taishukan. p. 39-52.
- KLOSA-KÜCKELHAUS, Annette; WOLFER, Sascha (2020). «Considerations on the acceptance of German neologisms from the 1990s». *International Journal of Lexicography*. p. 150-167.

- NAM, Kilim; LEE, Soojin; JUNG, Hae-Yun; CHOI, Jun (2016). «The life and death of neologisms. On what basis shall we include neologisms in dictionary?». A: MARGALITADZE, Tinatin; MELADZE, George (eds.). *Proceedings of the XVII EURALEX International Congress: Lexicography and Linguistic Diversity*. Tbilisi: Ivane Javakhishvili Tbilisi State University. 389-393.
- O'DONOVAN, Ruth; O'NEIL, Mary (2008). « A systematic approach to the selection of neologisms for inclusion in a large monolingual dictionary». A: *Proceedings of the 13th Euralex International Congress*. Barcelona. p. 571-579.
- SÁNCHEZ MANZANARES, Carmen (2013). «Valor neológico y criterios lexicográficos para la sanción y censura de neologismos en el diccionario general». *Sintagma* 25. p. 111-125.
- SHEIDLOWER, Jesse T. (1995). «Principles for the inclusion of new words in college dictionaries». *Dictionaries* 16. p. 33-44.

Notes

1. Aquest treball forma part del projecte de recerca finançat per l'Institut d'Estudis Catalans PRO2018-2021S04-FREIXA.
2. <https://www.upf.edu/web/nadic>
3. <http://www.iula.upf.edu/rec/neoxoc/>
4. <http://bobneo.upf.edu/>
5. En aquest cas, falta informació perquè encara no s'ha completat la cerca manual d'algunes obres.