

LEXICOGRAFIA CATALANA DIGITAL: PER UN CANVI DE PARADIGMA

Elisenda Bernal*

Resum

En aquest article s'analitzen els diccionaris en línia de llengua catalana, que s'avaluen des del punt de vista de la digitalitat i l'accés a les dades, amb l'objectiu de plantejar els reptes que ha d'assumir la lexicografia catalana de manera més imminent per oferir productes més orientats a l'usuari i a les seves necessitats.

Paraules clau: lexicografia; diccionaris; accessibilitat; digitalitat; usuaris.

DIGITAL CATALAN LEXICOGRAPHY: TOWARDS A PARADIGM SHIFT

Abstract

This paper analyses online Catalan dictionaries and assesses them according to their degree of digitality and the access to data they offer. The aim of this analysis is to lay out the challenges that Catalan lexicography must meet in order to offer products that are more closely oriented to users and their needs.

Keywords: lexicography; dictionaries; accessibility; digitality; users.

* Elisenda Bernal, doctora en lingüística per la Universitat Pompeu Fabra (UPF), professora agregada interina a la Facultat de Traducció de la UPF i membre del grup de recerca InfoLex i de l'Observatori de Neologia de la mateixa universitat. elisenda.bernal@upf.edu

Citació recomanada: Bernal, Elisenda. (2019). Lexicografia catalana digital: per un canvi de paradigma. *Revista de Llengua i Dret, Journal of Language and Law*, 71, 238-251. <https://doi.org/10.2436/rld.i71.2019.3273>

Sumari

1 Presentació

2 Diccionaris a la xarxa

2.1 Sobre la disponibilitat de diversos mètodes de cerca

2.1.1 Hipertextualitat

2.1.2 Identificació del lema i cerca incremental

2.1.3 Unitats polilexemàtiques

2.1.4 Entrades homonímiques i polisèmiques

2.2 Sobre la incorporació de material multimèdia

3 Conclusions: tanta digitalització i tan poca digitalitat

Referències bibliogràfiques

1 Presentació

El Baròmetre de l'ús del català a internet (WICCAC, 2018) situa el percentatge d'ús global del català en un 65,33% (dades del maig de 2018) i la llengua catalana és una de les deu llengües més actives de la xarxa. En aquest sentit, les notícies, doncs, són bones. I encara ho són més si atenem la presència de diccionaris de català disponibles a la xarxa: la situació no es pot titllar d'escassa, certament, però si es fa una mirada atenta i crítica als recursos lexicogràfics que hi ha, aquesta presència no compleix tots els requisits esperables i desitjables. Presentem breument a continuació els principals diccionaris de referència per a la llengua catalana disponibles a la xarxa, seguits d'una anàlisi de les característiques que creiem que hauria de tenir la lexicografia digital, sense entrar a comparar i analitzar les característiques de la macroestructura o la microestructura, més enllà del que sigui pertinent per analitzar-ne la digitalitat.

2 Diccionaris a la xarxa

La consulta de recursos lexicogràfics en línia ha crescut de manera exponencial, tot i que pocs usuaris volen pagar per fer-ho, cosa que ha fet caure molt significativament les vendes dels exemplars impresos, de manera molt més significativa que en altres sectors editorials (Dickson, 2018). Per al català, es compta amb dos grans diccionaris a la xarxa: per una banda, el *Gran diccionari de la llengua catalana* (GDLC) d'Enciclopèdia Catalana, publicat el 1998, tant en paper com en la seva versió en línia (www.diccionari.cat),¹ i el diccionari acadèmic de l'Institut d'Estudis Catalans (DIEC2, en línia a <http://dlc.iec.cat>), la segona edició del qual (2007), que va augmentar en més de 2.400 entrades en relació amb l'edició anterior de 1995,² es va publicar simultàniament en el portal de l'Institut, amb una interfície de consulta que millorava considerablement la de la primera edició i a la qual tornarem més endavant.³ Les diferències entre els dos diccionaris són fonamentalment el nombre d'entrades: mentre que el DIEC2 en té aproximadament 67.500, el GDLC n'incorpora 16.000 més, diferència que es reflecteix en l'aspecte de l'entrada, ja que les paraules coincidents tenen una marca visual (■), tal com es pot veure a l'entrada de *serigrafia*, enfront de *sostenibilitat*, que no hi és, o *thriller*, que, a més, presenta un asterisc per indicar que es tracta d'un manlleu:

▪ **serigrafia** *f f* GRÀF/TÈXT Procediment d'impressió i d'estampació de teixits en el qual el motiu és dibuixat en una tela de sedàs o en un teixit de seda muntat en un marc, i que hom aplica i tinta sobre el paper, el teixit, etc., que hom vol imprimir.

sostenibilitat *f* ECOL/ECON Conjunt de polítiques destinades a fer compatible el creixement econòmic i la preservació de la biodiversitat i evitar, en darrer terme, la degradació de la biosfera provocada per l'acció humana.

thriller* *m* CIN Mot emprat per a definir una pel·lícula de suspens, policíaca o de terror, que provoca una forta tensió emotiva en els espectadors.

Des de febrer de 2016, a més, el DIEC2 ofereix la possibilitat de consultar la resta de recursos institucionals. Així, quan s'hi busca un lema determinat, en el resultat, a més d'oferir la paraula corresponent, apareixen els botons d'accés a la resta de recursos lexicogràfics que s'han vinculat al diccionari normatiu:

- DDLC = *Diccionari descriptiu de la llengua catalana*, diccionari d'ús que forma part del *Diccionari del català contemporani*, projecte de l'IEC des de 1985. En línia a <http://dcc.iec.cat/ddlc/index.asp> des de 2005. El nombre d'articles redactats i accessibles és de 109.997 (gener de 2019).

1 Actualment també és possible consultar el GDLC des de la pàgina www.enciclopedia.cat, que, de fet, és l'espai on s'actualitza el diccionari sistemàticament des de 2015, tant des del punt de vista de la macroestructura com de la microestructura (per exemple, només s'ha actualitzat l'ortografia dels diacrítics a la versió de l'accés des de www.enciclopedia.cat; vegeu, per exemple, l'entrada del verb *donar*). Aquesta duplictat de versions genera diversos problemes que comentem al llarg de l'article.

2 Vegeu Ginebra (2007) per a una anàlisi comparativa de les dues edicions del diccionari acadèmic.

3 La primera edició del DIEC és de 1995, però la versió en línia no va estar disponible fins a finals d'abril de 2001.

- CTILC = *Corpus textual informatitzat de la llengua catalana*, projecte de l'IEC des de 1985. En línia a <http://ctilc.iec.cat> des de 2001; des del 2016 hi ha una versió simplificada de consulta d'accés des de <http://ctlc.iec.cat>.⁴ El nombre de lemes que inclou és de 148.410 (gener de 2019).
- DCVB = *Diccionari català – valencià – balear* d'A. M. Alcover i F. de B. Moll (1932-1962), que conté més de 160.000 entrades, que reuneix la llengua literària, medieval i moderna, amb equivalències en castellà i una gran informació sobre les pronúncies dialectals. En línia des de 2003.
- Sinònims = *Diccionari de sinònims* d'Albert Jané (3a edició, de 2015), membre de la Secció Filològica, amb 27.542 articles.
- CiT = Portal de Ciències i Terminologia, amb accés a 62 obres de lèxic especialitzat (que es corresponen amb 175.919 termes diferents).⁵ Permet la cerca global en el conjunt de tots els diccionaris o bé en cadascun d'ells. Si es tria aquesta opció, el motor de cerca permet obtenir resultats per cada àmbit de l'arbre de camp, per inicis de paraula, finals, per llengües d'origen, etc.

Ara bé, la informació que s'obté és desigual, si és que no es pot arribar a qualificar de desconcertant, ja que si, per exemple, se cerca *agrella*, el DDLC no mostra directament l'entrada, sinó que s'ha de fer un clic més per accedir a l'article; el CTILC empitjora el resultat, ja que demana que l'usuari torni a introduir el lema que se cerca, tal com també ocorre amb el CiT, que a més retorna com a resultats *sal d'agrella*, del *Diccionari enciclopèdic de medicina*, i *vinagrella*, del *Diccionari de ramaderia*; amb la cerca al diccionari de sinònims s'obté que "Hi ha 0 articles que contenen «agrella»", i és només al DCVB on la consulta és directa i enllaça a l'entrada corresponent.⁶

El DIEC2 pot consultar-se, així mateix, des de l'Optimot (<http://optimot.gencat.cat>), portal d'informació lingüística que ajuda a resoldre dubtes diversos sobre la llengua catalana, i que integra, a més del diccionari acadèmic, diverses obres de referència: els diccionaris català-castellà i castellà-català d'Enciclopèdia Catalana, els verbs conjugats, l'*Ortografia catalana* de l'Institut, els diccionaris terminològics del Termcat, les fitxes de l'Optimot (en les quals es resolen dubtes lingüístics freqüents), la col·lecció "Criteris Lingüístics" de la Direcció General de Política Lingüística, i els nomenclàtors de toponímia de Catalunya i la Catalunya del Nord. No és, doncs, una oferta de mínims, al contrari. Un dels avantatges de fer una consulta des de l'Optimot i no des de cadascuna de les fonts concretes és que, si no es restringeix a priori per algun motiu, els resultats que s'obtenen recullen totes les fonts en què la paraula buscada apareix, ja sigui com a lema o com a part de la definició en els diccionaris, o que aparegui citat en alguna explicació com a exemple en el cas de les fitxes o de l'*Ortografia*, cosa que explica el nombre elevat de resultats, com es pot veure en el cas de *sabó*:

<i>Diccionari de la llengua catalana</i> (28)
<i>Diccionari castellà-català</i> (9)
<i>Diccionari català-castellà</i> (8)
Fitxes de l'Optimot (2)
<i>Ortografia catalana</i> (5)
Diccionaris terminològics (12)

Taula 1. Resultats per a la cerca de *sabó* a l'Optimot

Encara cal afegir altres recursos de naturalesa diversa a la llista:⁷

4 L'Institut d'Estudis Catalans preveu que l'estiu de 2019 aquests accessos seran substituïts per una interfície nova més funcional que, a més, permetrà consultar el corpus ampliat amb els textos que s'estan incorporant.

5 Dades de desembre de 2018.

6 A més de millorar els enllaços i que no calgui tornar a escriure el lema ni oferir-ne d'aproximat, es podria plantejar que la consulta només estigués activa quan el recurs en qüestió inclou efectivament la paraula que se cerca, recorrent a un to de gris més fosc, per exemple.

7 Aquesta llista no pretén ser exhaustiva, sinó que només recull una mostra de diversos recursos, tant institucionals com personals, que exemplifiquen el ventall de material lexicogràfic en línia de què disposa el català.

- Acadèmia Valenciana de la Llengua (2014). *Diccionari normatiu valencià*. En línia a <http://www.avl.gva.es/lexicval> des de 2014.⁸
- *Diccionari de la llengua catalana multilingüe*. Barcelona: Enciclopèdia Catalana, 1997. Diccionari multilingüe que permet la consulta en cinc llengües: català, castellà, anglès, francès i alemany. En línia a <http://www.multilingue.cat>.
- *Diccionari enciclopèdic de medicina*. Barcelona: Enciclopèdia Catalana, 1990. Diccionari terminològic de medicina, la segona edició del qual (2012) s'ha integrat en el CiT, amb la qual cosa ha ampliat les opcions de cerca de manera considerable. En línia a <http://www.medic.cat> o <http://cit.iec.cat/obres.asp?obra=23>.
- *Didac. Diccionari de català*. Barcelona: Enciclopèdia Catalana, 1995. Diccionari escolar en línia a <http://www.dicdidac.cat> des de 2001.
- Espinal, M. Teresa (2004). *Diccionari de sinònims de frases fetes*. Bellaterra; València; Barcelona: Servei de Publicacions de la Universitat Autònoma de Barcelona; Publicacions de la Universitat de València; Publicacions de l'Abadia de Montserrat. En línia a <http://dsff.uab.cat> des de 2018, a partir de la segona edició del diccionari (2006), i que inclou 5.500 entrades conceptuals i 15.500 expressions lexicalitzades.
- Fundació Wikimedia (2004–). *Viccionari*, en construcció. En línia a <https://ca.wiktionary.org/wiki/Portada>. Aquest diccionari es descriu com un “diccionari universal lliure en desenvolupament continu [...]” i que pretén “recollir totes les paraules del món i explicar-les en català”, característiques que són l'avantsala d'incloure paraules molt diverses i que són fàcilment qüestionables des del punt de vista de la lexicografia. El desembre de 2018 comptava amb 374.211 entrades de diccionari, nombre que es deu al fet que incorpora noms propis, així com entrades independents per a les formes flexionades de noms i adjectius.
- Observatori de Neologia (2014–). *Neolosfera. Cada dia, una paraula nova*. Barcelona: Universitat Pompeu Fabra, en construcció. Blog de l'Observatori de Neologia en què cada dia laborable es publica un neologisme lexicogràfic. Si bé no és pròpiament un diccionari, des de la pestanya “Arxiu de neologismes”, que s'actualitza setmanalment, es va construir a poc a poc una llista amb els neologismes publicats amb els hipervincles degudament enllaçats a les entrades corresponents en què es troba l'explicació i la definició. En línia a <https://neolosfera.wordpress.com> des de 2014.
- Oliveras, S. (dir.) (2002). *Diccionari pràctic i complementari de la llengua catalana*. Barcelona: La Busca Edicions. En línia a <http://www.dicpc.cat> des de 2007. Inclou apartats gramaticals i presenta llistes de paraules que es consideren barbarismes amb els equivalents normatius.
- Portal *diccionaris.cat*, de l'editorial Larousse, que integra la tercera edició del *Diccionari de llengua catalana manual* (2015), un diccionari de sinònims, i els diccionaris bilingües de castellà, anglès i francès en combinació amb el català. En línia a <http://diccionaris.cat> des de 2010.
- Termcat. Centre de Terminologia Catalana (1985–). Coordina l'activitat terminològica en llengua catalana, creant productes terminològics i normalitzant neologismes. Des del Cercaterm (<http://www.termcat.cat/es/Cercaterm/Fitxes>) s'accedeix al conjunt de més de 230.000 fitxes de termes i criteris que s'actualitzen regularment. Està proveït d'una modalitat bàsica de consulta per aproximació, de manera que obté com a resultats tant la coincidència plena com la variació (fins al 79%), així

⁸ Inclouem el diccionari de l'Acadèmia Valenciana de la Llengua, malgrat el conflicte de competències que es genera en alguns casos davant de determinades propostes que no coincideixen amb les de l'Institut d'Estudis Catalans, l'autoritat lingüística oficial, reconeguda com a tal en el Reial decret 3118/1976, de 26 de novembre, ratificada per la Llei 8/1991, de 3 de maig, del Parlament de Catalunya. Amb la creació de l'Acadèmia Valenciana de la Llengua mitjançant la Llei 7/1998, de 16 de setembre, de la Generalitat Valenciana, tot i que en cap moment es diu que el valencià sigui una llengua diferent, més enllà dels problemes polítics, l'autoritat legal reconeguda que té genera problemes per als parlants de la Comunitat Valenciana, que es troben amb el dilema d'haver de triar entre seguir unes indicacions o unes altres quan s'ofereixen solucions diferents. Vegeu Costa (2008) per a una reflexió sobre aquesta dualitat de normes.

com d'una modalitat avançada que permet escollir entre denominació, definició i nota, buscar per llengua, categoria gramatical i àrea temàtica. En línia des de 2000.

- Vilaseca, Oriol (2005). *Diccionari invers de la llengua catalana*. En línia a <http://dilec.org>.

Només amb aquestes dades i sense entrar gaire en detall encara, és clar que l'oferta lexicogràfica en llengua catalana a la xarxa sembla que gaudeix de bona salut quant a cobertura i diversitat de fonts: disposem de diccionaris monolingües, bilingües i multilingües, descriptius i prescriptius, generals i especialitzats, un diccionari escolar, i diccionaris que podem considerar complementaris, com són el *Diccionari invers* o la Neolosfera, a més del *Diccionari pràctic i complementari de la llengua catalana*. Ara bé, aquests diccionaris a la xarxa són productes *realment* digitals?⁹ Per respondre aquesta pregunta, serveix per començar la definició d'*hiperdiccionari* d'Abramson (1995: 10):

“A hyperdictionary is a comprehensive dictionary designed to take advantage of the special characteristics of electronic media: abundant storage space; the ability to create interlocking databases; availability of multiple methods to search for a given item; and multimedia capabilities that allow visual and oral as well as textual representations. Formally the hyperdictionary is defined below:

A relational and deductive database containing the words of a language or languages, together with an open-ended set of access and display methods so as to present at least their orthography, pronunciation (voiced by a native speaker), signification, part of speech, and use, their history, synonyms, homonyms, antonyms, derivation, relationships to one another, and any other aspect of the words that may be necessary for reference, teaching or study purposes. Additional information about language or languages, including grammar, morphology, semantics, pragmatics, machine tractable representations, etc., as well as relevant information concerning geography, names, literature, society, culture, history and so on, is not excluded from the database.”

De la definició d'Abramson destaca el fet que un hiperdiccionari permet interconnectar de manera àgil tots els elements que s'hi inclouen en forma d'una xarxa interactiva, i no de taula o de llista, de manera circular, en què no hi ha ni principi ni fi, de manera que l'usuari pot navegar pels diversos blocs d'informació i crear el seu propi camí. D'altra banda, però, tal com recull Lew (2012), no s'ha d'oblidar que, perquè la consulta a un diccionari sigui efectiva, al marge del suport en què estigui editat, ha de ser capaç de respondre les necessitats dels usuaris, en un temps raonablement curt i amb el grau de detall requerit, i, concretament, en el cas dels diccionaris digitals, cal afegir la importància de l'accés a les dades i el rol que hi juguen els elements multimèdia. Així doncs, tenint en compte tant la definició d'Abramson (1995) com les observacions de Lew (2012), analitzem a continuació si en els recursos lexicogràfics de què disposa el català realment es compleixen les característiques dels mitjans digitals, quant als paràmetres següents:

- disponibilitat de diversos mètodes de cerca
 - hipertextualitat
 - identificació del lema
 - unitats polilexemàtiques
 - entrades homonímiques i polisèmiques
- incorporació de material multimèdia

En l'apartat següent analitzem aquests paràmetres en els diccionaris esmentats, centrant-nos especialment en els diccionaris generals.

⁹ És freqüent trobar altres denominacions per referir-se als diccionaris digitals, com ara *diccionaris* electrònics o *diccionaris en línia*, si bé no són termes sinònims. Per a una discussió més detallada sobre el tema, vegeu Gelpí (2003).

2.1 Sobre la disponibilitat de diversos mètodes de cerca

Tots els diccionaris incorporen, lògicament, la cerca per lemes, però pocs van més enllà: el cercador de l'Optimot és el més ampli i, a més d'oferir els resultats en els quals apareix la paraula consultada, els resultats apareixen jerarquitats, tot i que no de manera explícita. Així, en primer lloc, apareixen els resultats en què la paraula coincideix amb el lema d'algun dels diccionaris, després els diccionaris en què la paraula apareix en la definició, i, finalment, en les entrades en les quals apareix com a exemple, tal com es recull a la taula 1 per al cas de *xantatge*:

com a lema ¹⁰	DIEC2 Diccionaris terminològics del Termcat <i>Diccionari català – castellà</i>
en la definició	DIEC2 (<i>s. v. xantatgista</i>) <i>Diccionari castellà – català (s. v. chantaje)</i> Diccionaris terminològics del Termcat (<i>s. v. extorsió sexual, porno venjatiu, sextorsió</i>)
en els exemples	<i>Diccionari castellà – català (s. v. malvat -ada)</i>
en els equivalents	<i>Diccionari castellà – català (s. v. achaque)</i>

Taula 2. Resultats per a la cerca de *xantatge* a l'Optimot

De la mateixa manera, la nova versió del GDLC, accessible des de www.enciclopedia.cat, ofereix resultats similars, que inclouen, de manera jerarquitada, en primer lloc, l'entrada o entrades que coincideixen exactament amb la paraula que s'ha buscat, i a continuació les entrades on apareix la paraula en altres camps de l'article. Així, per a *testar* 'fer testament', s'obtenen els resultats següents:¹¹

com a lema	<i>s. v. testar</i> ¹
com a remissió	<i>s. v. testar</i> ²
en la definició	<i>s. v. lligítima</i> <i>s. v. preterició</i> <i>s. v. substitució</i> (subentrada <i>substitució exemplar, substitució pupil·lar</i>) <i>s. v. llibertat</i> (subentrada <i>llibertat de testar</i>)
en els exemples	<i>s. v. capaç</i>
en l'etimologia	<i>s. v. testador testadora</i> <i>s. v. testes</i>
en els homòfons	<i>s. v. tastar</i>

Taula 3. Resultats per a la cerca de *testar* al GDLC

Per la seva banda, el DIEC2 permet unes possibilitats de cerca insòlites per a un recurs gratuït sota l'opció "Consulta avançada": per entrades, subentrades, definicions, exemples, categories gramaticals, marques valoratives, llengua d'origen o àrea temàtica, i amb la possibilitat d'incloure condicions múltiples de cerca per a cada cas: a més de la consulta de la forma coincident, que és la que està establerta per defecte, permet activar els paràmetres "començada per", "acabada en", "en qualsevol posició", "no començada per", "no acabada en", "que no contingui", amb la qual cosa es pot accedir a molta informació diversa d'una manera molt àgil.¹²

¹⁰ En el cas que el lema de consulta formi part de termes sintagmàtics, apareixen llistats després dels resultats del diccionari general. Així, de *potestat* es recull *potestat administrativa, potestat expropiadora, potestat jurisdiccional, potestat parental*, etc.

¹¹ A més, els usuaris registrats poden accedir des de cada resultat a informació complementària (*Diccionari de sinònims Franquesa* i conjugació en el cas dels verbs) i als equivalents en alemany, anglès, castellà, francès i italià), només quan hi ha resultats. Així, *minorista* només dona opció a consultar els equivalents en castellà i francès, però *llegar* dona l'oportunitat de desplegar tota la informació complementària i tots els equivalents.

¹² Les consultes d'accés al DDLC, en el moment de redacció d'aquest article (desembre de 2018), encara són limitades, però està previst que s'oferirà una interfície oberta i amb diverses opcions de consulta sobre les entrades i sobre els diversos elements estructurals del diccionari.

Finalment, la Neolosfera, que és un blog, inclou un motor de cerca en què és possible buscar qualsevol paraula que s'hagi publicat, ja sigui com a entrada o que aparegui en els contextos o la definició, si bé no va més enllà de la constatació.

2.1.1 Hipertextualitat

Encara que la llista de diccionaris de llengua catalana és ingent, la hipertextualitat sembla quedar-ne al marge, malgrat que potser és la característica més prominent dels diccionaris digitals, tot i que no és al caire del buit absolut, ja que el GDLC, així com també el *Didac*¹³ i el *Diccionari normatiu valencià* (en endavant, DNV), sí que són hipertextos, i força aconseguits: qualsevol paraula que apareix en la definició o en els exemples, encara que aparegui flexionada, pot ser clicada i navegar pel diccionari molt fàcilment. No obstant això, amb el GDLC es dona una situació curiosa: la consulta des de www.diccionari.cat és, efectivament, un hipertext, que genera soroll en alguns casos, com ara que incorpora totes les paraules que hi apareixen, de manera que les abreviatures de l'etimologia i la codificació temàtica i lingüística es tracten de la mateixa manera que la resta de paraules, amb la qual cosa l'usuari rep informació parcial i inadequada. Així, si es consulta *tatà* 'cavall, en llenguatge infantil', per exemple, si es clica sobre la marca *pop* que encapçala la definició, obre un calaix en què s'ha de triar entre dues paraules homònimes, els significats de les quals disten molt del que realment significa en aquest context, ja que confon l'animal i l'estil musical o artístic amb l'abreviació d'una marca valorativa associada al registre popular.

La funció hipertextual s'aplica també a l'etimologia, sobre la qual qualsevol clic és inútil, ja que els ètims no tenen entrada, si no és que coincideixen plenament amb una paraula catalana, com passa amb *olor* (del llatí *olor*) o *barba* (del llatí *barba*). És més compromès el que passa quan es clica sobre una paraula de la definició iniciada per vocal i que està precedida per un article o una preposició apocopats: en alguns casos, no reconeix cap de les dues parts, de manera que envia a una llista de paraules pròximes, amb la particularitat que és una seqüència de paraules bastant llunyana a la del substantiu que es busca, com succeeix amb *marmessor* en clicar sobre *d'executar*:

Vegeu la llista de paraules més properes alfabèticament a d'executar :	
devorar	dextrinació
devot	dextrinar
devotament	dextro-
devuit	dextrocàrdia
dexametasona	dextrogir
dexio-	dextrorotatori
dextrà	dextros
dextrat	dextrosa
-dextre	dg
dextrina	dg.

Taula 4. Proposta de paraules pròximes a *d'executar* (GDLC)

Val a dir, però, que en la versió actualitzada del diccionari accessible des de www.encyclopedia.cat, aquests problemes s'han resolt, ja que les marques valoratives no apareixen abreujades sinó totalment desplegades, amb una tipografia diferent: en color gris clar, diferenciat clarament del negre de la definició, i en versaleta en el cas de les marques temàtiques i en cursiva en cas de les marques lingüístiques (vegeu, per exemple, *bici*).

2.1.2 Identificació del lema i cerca incremental

¹³ Tanmateix, cal fer notar que el *Diccionari enciclopèdic de medicina* i el *Diccionari de la llengua catalana multilingüe*, de la mateixa editorial, no són hipertextos. En el cas del primer, és lògic, atès que es tracta d'un diccionari terminològic i no totes les paraules que s'utilitzen en la definició hi tenen entrada. En el segon cas, és menys comprensible, ja que seria interessant poder navegar entre les diverses llengües d'una manera més àgil.

La consulta per defecte en qualsevol diccionari és la consulta per lema. Ara bé, a partir del lema, la motivació última de l'usuari pot ser molt diversa. Lamentablement, en català no disposem d'estudis sobre les cerques que realitza un usuari en un diccionari general de català, però l'estudi de Ferret (2014) sobre les consultes més freqüents fetes a l'Optimot ofereix algunes pistes:¹⁴

Categoria	2012	2013	Total
Verbs	32,2	33,2	32,7
Ortografia i accentuació	24,2	26	25,1
Castellanismes	15,2	15,2	15,2
Lèxic	7,8	7,6	7,7
Conjuncions	4,6	4,8	4,7
Preposicions	4,6	4,4	4,5
Pronoms	1,2	1,8	1,5
Adverbis	1,2	0,8	1
Adjectius	1	1,2	1,1
Gènere i nombre	1,2	0,8	1

Taula 5. Percentatge de consultes a l'Optimot

Atenent aquests resultats, no exempts de problemes en la classificació de consultes, i també basant-nos parcialment en els estudis d'Azorín (2010) i Cassany (2016), realitzats sobre enquestes i entrevistes, respectivament, a alumnes de secundària i batxillerat, és molt probable que les consultes es vertebrin al voltant del significat, l'ortografia, la construcció sintàctica, la categoria (o millor, subcategoria) gramatical i la fraseologia.¹⁵

I l'ortografia és aquí un problema: les interferències ortogràfiques amb el castellà són constants, amb la qual cosa el primer problema és que si no se sap com s'escriu una paraula es pot no obtenir cap resultat. És cert que el DIEC2 i el DNV permeten certa relaxació ortogràfica: en accedir al diccionari es pot buscar una paraula sense cap accent ni dièresi i obtenir la resposta desitjada: així, una consulta com *pingui* condueix a l'entrada de *pingüi*. Ara bé, si es busca *consecuencia* en el DIEC2 s'obté com a resultat el missatge "No s'ha trobat cap entrada coincident amb els criteris de cerca". El resultat és igualment negatiu en el GDLC, però a la pantalla apareixen, si més no, les paraules immediatament anteriors i posteriors a la paraula que s'ha buscat i, on, amb sort, pot aparèixer el mot consultat, com succeeix amb *pingui*, però que no ocorre amb una consulta com *babós*, on els resultats que ofereix el GDLC queden molt lluny de la forma que es vol realment consultar, *bavós*.¹⁶ Només el DNV resol parcialment aquest problema: si bé la recerca *consecuencia* obté un "No hi ha resultats" com a resposta, la de *babós* és molt més afínada, ja que ofereix directament com a resultats tant *barbós* com *bavós*. En canvi, la consulta de *palidesa* no preveu l'opció del dígraf i retorna propostes molt allunyades (*calidesa, polidesa, validesa*) de *pal·lidesa*.

Cal afegir encara un altre problema, tot i que afecta un percentatge relativament petit d'entrades: com hem dit, el DIEC2 i el DNV tenen l'avantatge que es pot prescindir de l'accentuació en fer la cerca, però aquest avantatge s'esvaeix davant de parells com *bestia – bèstia* o *sol – sòl*: mentre que en el DIEC2 queda minimitzada, perquè el resultat ofereix les dues opcions i l'usuari ha de clicar en cadascuna d'elles per confirmar quina és la que volia consultar, en el DNV només s'obté la forma sense accent.

¹⁴ Reproduïm només els percentatges de les consultes més freqüents amb resultat $\geq 1\%$.

¹⁵ No cal dir que tant el castellà com el català són lluny encara d'estudis com els de Heid i Zimmerman (2012), ja sobre diccionaris en línia, sobre el registre de les pulsacions en diverses interfícies de diccionari, o de treballs que utilitzen programes de seguiment de la mirada per descobrir quines àrees de la pantalla són les que efectivament es consulten, en quin ordre i durant quant de temps (vegeu, per exemple, Tono, 2011, Kaneta, 2011, i Simonsen, 2011). Són necessaris estudis d'aquesta mena per saber què és el que necessiten els usuaris per tal de millorar les prestacions dels diccionaris.

¹⁶ Aquesta observació val per a la consulta des de www.diccionari.cat. A www.encyclopedia.cat, el resultat que ofereix és "Potser volíeu dir *consecuencias*", que envia seguidament a "Potser volíeu dir *consecutius*", on es despleguen 29 resultats en què apareix, sense que apareguin adequadament jerarquitats, ja que l'adjectiu *consecutiu consecutiva* apareix en tercer lloc, després de les entrades de *codó* i *patèrgia*. En canvi, la cerca de *pingui* a www.encyclopedia.cat, tot i que proposa que "Potser volíeu dir *pinguis*" (on s'inicien, de nou, més propostes: *pinguinus, pingüicula, unguiculat, ungulat, ungulats*), dona també el resultat directe de tres entrades: *pingüi, diacrític* i *dièresi*, en què amb la primera sí que s'obté la informació que es busca.

Finalment, en aquest apartat, cal esmentar les possibilitats de consulta de formes. El DIEC2 permet la consulta de substantius i adjectius amb tota la flexió desplegada: masculí, femení, singular i plural. Així, si s'introdueix *vampires* en el camp d'entrada, la cerca ens porta a l'entrada de *vampir vampira*; o en el cas de *grogues* porta a *groc groga*. Com a contrapartida, aquesta funció també genera força soroll en altres casos, tal com es pot veure a la taula 6:

cerca	resultat
<i>llavors</i> 'aleshores'	<i>llavor</i> 'germen' <i>llavors</i> 'aleshores'
<i>manera</i> 'mode, manera'	<i>maner -a</i> 'manejable' <i>manera</i> 'mode, manera'
<i>segons</i> 'guardant tal o tal direcció'	<i>segó</i> 'pellofa de cereals' <i>segon</i> 'unitat de temps' <i>segon -a</i> 'que segueix immediatament el primer' <i>segons</i> 'guardant tal o tal direcció'
<i>taula</i> 'moble'	<i>taül -a</i> 'pocavergonya' <i>taula</i> 'moble'

Taula 6. Resultats sorollosos del DIEC2

Aquest soroll seria fàcilment eliminable si s'incorporés l'opció, ara disponible només des de la consulta avançada, denominada "puntuació gràfica", de manera que l'usuari podria obtenir la informació desitjada. Incomprensiblement, aquesta funció, tanmateix, no s'aplica al paradigma verbal: si busquem *comprenia* o *desfés*, el resultat és negatiu, lematització que sí que té integrada el DNV i que retorna com a resultats *comprendre* i *desfer*, respectivament. No obstant això, l'ortografia establerta per l'AVL no coincideix totalment amb la de l'IEC, de manera que si bé la forma del participi *estès* permet arribar a *estendre*, no ocorre el mateix amb *malmès* o *comprès*, ja que l'ortografia acadèmica valenciana només els recull amb accent agut.

2.1.3 Unitats polilexemàtiques¹⁷

Una de les dificultats amb les quals s'ha d'enfrontar qualsevol usuari que s'acosta al diccionari és el moment temut de cercar una unitat polilexemàtica i decidir per quin dels lexemes que la conformen ha de plantejar la consulta: *peu de cabra* apareix sota l'entrada de *peu* (DIEC2, GDLC) i *gas noble* o *gas natural* són subentrades de *gas* (DIEC2, GDLC). Ara bé, *gas mostassa* no apareix en el DIEC2 i en el GDLC apareix sota el lema de *mostassa*. Així mateix, per trobar en el DIEC2 *efecte d'hivernacle* cal fer la consulta a *efecte*, mentre que en el GDLC només el trobarem dins *hivernacle* (a més que la unitat perd la preposició a favor d'*efecte hivernacle*), i així com *novel·la breu* és una subentrada de *novel·la* en tots dos diccionaris, *novel·la rosa* apareix dins *novel·la* en el DIEC2, mentre que es recull com a subentrada de *rosa* en el GDLC (tot i que *premsa del cor* (o *rosa*) és dins de *premsa*).

Sense cap mena de dubte, aquest aspecte és un galimaties fins i tot per als especialistes en llengua, que no aconsegueixen trobar la clau per saber on cal buscar. Seria possible, doncs, plantejar una consulta més enllà del lema, que no deixa de ser una restricció de consulta heretada de la consulta tradicional del diccionari en paper? La resposta és afirmativa, ja que l'Optimot la permet, tal com s'ha mostrat anteriorment en els mètodes de cerca, però és l'únic dels recursos que la té integrada.¹⁸

¹⁷ Vegeu Bach i Matamala (2004) per a les dificultats que suposa la cerca en diccionaris en el cas de connectors i conjuncions.

¹⁸ Cal dir que la consulta del GDLC des de www.encyclopedia.cat permet introduir les cerques entre cometes o utilitzar els signes de suma (+) i resta (-) per afinar la cerca. En el cas dels exemples, tant la cerca entre cometes o amb l'ús de la suma retornen adequadament el resultat. En cas que s'introdueixi sense les cometes, els resultats es multipliquen, tot i que sí que apareixen jerarquitzats i la primera entrada que ofereix és la que conté el sintagma cercat.

2.1.4 Entrades homonímiques i polisèmiques

L'homonímia i la polisèmia són dos fenòmens semàntics no sempre fàcils de delimitar, però de la concepció que se'n té depèn la representació del significat que s'ofereix en els diccionaris: en lexicografia, més enllà de discutir sobre els seus límits, s'adopta un criteri metodològic per adscriure un significat concret dins una entrada o bé per diferenciar-lo en una entrada independent quan l'etimologia o la diferenciació significativa o funcional ho recomana (García Platero, 2004): així, per exemple *flamenc* té dues entrades homògrafes en el DIEC2 (una per a l'ocell, i l'altra per al natural de Flandes i el caràcter popular andalús), mentre que el DNV, el GDLC o el *Diccionari de la llengua catalana multilingüe* reuneixen els tres significats en una sola entrada.

En la lexicografia tradicional els dos fenòmens suposen un cost d'espai considerable, problema que es minimitza en el mitjà digital, si bé, de nou, encara és subsidiari del predecessor en paper. Així, pel que fa a la polisèmia, encara que el DIEC2 fa un pas de gegant en comparació amb la primera edició quant a la presentació, el GDLC la millora amb una tipografia més gran i amb concessions ortotipogràfiques que ajuden l'usuari: sagnats, interlineat generós i lletra més gran, tot i que no deixa de ser un sistema complex per als usuaris que, en les entrades llargues, han d'anar passant d'un significat a un altre sense saber exactament el criteri que s'amaga darrere de l'organització microestructural.

Quant a l'homonímia, el GDLC ofereix resultats diferents segons el lloc des del qual es consulta: si es consulta des de www.diccionari.cat les entrades apareixen correlativament una sota l'altra, de manera que només cal desplaçar-se per la pantalla en una sola pàgina. En canvi, des de www.encyclopedia.cat, de la mateixa manera que els diccionaris acadèmics, es proporcionen les entrades d'una en una, de manera que l'usuari ha d'anar-les seleccionant fins que troba la que realment busca, com, per exemple, ocorre amb *pinta*, que té tres entrades diferenciades: una per a l'estri, una altra per a l'aspecte i l'última per a la mesura de capacitat.

Encara sobre la polisèmia, cal destacar que el DIEC2 i el DNV inclouen enllaços per a les remissions que coincideixen amb subaccepcions: així, per exemple, en el DIEC2 l'accepció 5 d'*abisme* remet a 'Infern 1 i 2.1', o el DNV recull en la quarta accepció de *circumstància* la subentrada *cara de circumstància* que remet a *cara*¹ 17.¹⁹ En tots dos casos, en clicar sobre els números, el diccionari salta a l'accepció concreta i no mostra més informació. En aquest sentit, els diccionaris acadèmics afinen molt més que el GDLC que, en no discriminar la informació de la microestructura i tractar-la tota com un únic hipertext a www.diccionari.cat, no aconsegueix conduir l'usuari en aquests casos: així, a *faltar*, la primera accepció del qual remet a *mancar* 2.1 i 2.2, només podem arribar a l'entrada del verb i no a les accepcions concretes; en aquest cas, en què només hi ha quatre accepcions no és un problema, però lògicament pot ser molt més complex.

2.2 Sobre la incorporació de material multimèdia

Només dos diccionaris, el DCVB i el *Didac*, incorporen material visual, si bé és estàtic i reproduïx les il·lustracions de les versions respectives en paper. En el cas del DCVB es tracta de dibuixos, realitzats per Francesc de Borja Moll, d'algunes plantes i objectes típics, com els que es troben a *baldufa*, *forca* o *grèvol*, per exemple. Ara bé, així com en el diccionari imprès la imatge apareix al costat de l'entrada, en la versió en línia s'indica que l'entrada en qüestió inclou una il·lustració mitjançant una icona d'una càmera fotogràfica que, en clicar-hi al damunt, fa aparèixer una finestra amb el dibuix.

Per la seva banda, el *Didac* inclou 40 làmines temàtiques sobre temes com la casa, el cos humà, el carrer, la música, etc., a més d'incorporar algunes il·lustracions puntuals (400 sobre 15.000 entrades, un 2,66%), com ocorre a *descabdellar*, *dinosaure* o *tripode*, que alleugereixen la quantitat de lletra i el fan més amè per al públic objectiu al qual s'adreça el diccionari.

Cal fer un esment especial en aquest apartat al diccionari de llengua de signes catalana (LSC) *Mira què dic. Diccionari multimèdia de signes de Catalunya*, publicat pel Departament d'Educació – Universitat de Barcelona (l'última versió del qual és de 2010), en línia a <http://www.edu365.cat/mqd>, que no analitzem

¹⁹ També el DDLC incorpora la mateixa funcionalitat, amb la diferència que en clicar sobre el número s'obre una finestra nova amb l'accepció concreta sense que es canviï l'entrada que s'estava consultant en primer lloc.

perquè es tracta d'una altra llengua, però que cal tenir en compte perquè és una mostra de diccionari que incorpora, necessàriament, material audiovisual de manera sistemàtica per a totes les entrades.²⁰

Fora d'aquests casos, la integració de material audiovisual o multimèdia és inexistent en la lexicografia catalana en línia.

3 Conclusions: tanta digitalització i tan poca digitalitat

Arribats a aquest punt, la revisió dels recursos presentats té una valoració agredolça: es pot afirmar que la llengua catalana es troba en una posició privilegiada quant a l'oferta de productes lexicogràfics en línia, gràcies a l'esforç d'iniciatives públiques i privades, de manera que accedir-hi és molt més efectiu i eficient en termes de rapidesa per a l'usuari que els diccionaris tradicionals en paper. Tanmateix, el camí a la digitalitat real està només iniciat: tots els diccionaris s'han concebut encara d'una manera tradicional i el traspàs a la interfície digital no ha anat, en la majoria dels casos, més enllà de la hipertextualització, parcial o total (en alguns casos, fins i tot excessiva) del text.

El mitjà digital permet solucions imaginatives i permet albirar en un futur, que no hauria de ser gaire llunyà, un disseny més orientat als usos de consulta més habituals dels diccionaris. Per assolir-ho, certament, calen estudis d'ús diversos, però es poden apuntar algunes funcions que podrien incorporar-se (a més de la cerca integral que ja incorporen alguns dels recursos i que no només busquen al camp del lema):

- Cerca per reconeixement de veu, que permetria arribar fàcilment a grafies que tradicionalment es resisteixen entre els usuaris de català, com ara les de *dansa*, *bolcar* o *harmonia*.
- Cerca incremental, que permetria que en anar teclejant la paraula anessin apareixent les entrades, amb la qual cosa es podria minimitzar l'error ortogràfic.
- Per a les entrades polisèmiques, inclusió d'etiquetes semàntiques (amb una tipografia diferenciada, per orientar la lectura i la identificació de significats) agrupant accepcions relacionades.
- Possibilitat de mostrar la informació per nivells, de manera que l'usuari decideixi si vol veure l'entrada completa, totalment desplegada, o bé parcial (definicions, exemples, etc.).

I encara es podrien afegir altres funcionalitats, com ara la d'interactuar amb altres recursos d'internet que possibiliti treballar amb dades extralèxicogràfiques i altres fonts d'informació, especialment orientada a l'àmbit de l'ensenyament-aprenentatge de llengües, o la d'integrar l'accés a aclariments multimèdia, ja sigui amb representacions animades o vídeos (Sanmarco, 2008).

El repte és, doncs, que els pròxims diccionaris es planifiquin i es construeixin considerant-los des del primer moment productes digitals i no merament digitalitzats, que, més enllà d'agilitzar l'actualització de la informació que ofereixen, garantint, així, la seva vigència il·limitada, adequin les funcions que s'ofereixen al tipus de diccionari i d'usuari, i s'orientin cap a un canvi de paradigma: del diccionari estàtic i rígid al diccionari dinàmic i flexible. Sens dubte, s'han obert nous camins per a la lexicografia i cal avançar-hi maximitzant totes les seves fortaleses que, innegablement, en són moltes. D'aquesta manera, com ja apuntava Dodd (1994), el diccionari podrà deixar de ser merament un recurs per poder ser un servei.

²⁰ No obstant això, com a diccionari de llengua de signes, no deixa de ser subsidiari d'un diccionari d'una llengua oral, ja que l'ordenació segueix els principis de les llengües orals, atès que els primers diccionaris de llengües de signes, de la dècada dels anys seixanta del segle xx, tenien com a objectiu principal donar a la família i al professorat un llibre de referència que facilités la comunicació amb els nens sords. Actualment, des de l'Institut d'Estudis Catalans s'està treballant en un projecte per crear una base de dades que serveixi per a l'elaboració de materials lexicogràfics, basats totalment en els avantatges del format digital, de manera que es possibilitarà que la consulta sigui, a més de la glossa convencional de la llengua oral, mitjançant paràmetres fonològics: la configuració (forma que adopten les mans i els dits mentre s'articula el signe), la localització (lloc on s'articula), el moviment (acció de les mans) i l'orientació (direcció del palmell de la mà). Per a més detalls, vegeu Barberà i Ribera (2010).

Referències bibliogràfiques

- Abramson, Harvey. (1995). What is a hyperdictionary? *TermNet News*, 49, 7-13.
- Azorín, Dolores. (2010). Perspectivas en la investigación sobre el uso del diccionario en el ámbito docente. A: Elisenda Bernal, Sergi Torner i Janet DeCesaris (ed.), *Estudis de lexicografia 2003-2005* (p. 179-202). Barcelona: Institut Universitari de Lingüística Aplicada – Universitat Pompeu Fabra.
- Bach, Carme, i Matamala, Anna. (2004). La pluriaccessibilitat com a eix de la lexicografia hipertextual. A: Paz Battaner i Janet DeCesaris (ed.), *De lexicografia. Actes del I Symposium Internacional de Lexicografia* (p. 199-216). Barcelona: Institut Universitari de Lingüística Aplicada – Universitat Pompeu Fabra.
- Barberà, Gemma i Ribera, Eulàlia. (2010). Lexicografia de les llengües de signes i l'estat de la lexicografia en la LSC. A: Joan Martí i Josep M. Mestres (ed.), *Les llengües de signes com a llengües minoritàries: perspectives lingüístiques, socials i polítiques* (p. 207-237). Barcelona: Institut d'Estudis Catalans.
- Cassany, Daniel. (2016). Recursos lingüístics en línia: contextos, pràctiques y redes. *Signos, número extraordinari 1*, 7-29.
- Costa, Joan. (2008). Qui codifica estandarditza. A: *Jornades de la Secció Filològica de l'Institut d'Estudis Catalans a Castelló de la Plana en homenatge a les Normes de Castelló* (p. 87-99). Barcelona: Institut d'Estudis Catalans.
- Dickson, Andrew. (23 de febrer de 2018). Inside the OED: can the world biggest dictionary survive the internet. *The Guardian*.
- Dodd, W. Steven. (1994). La lexicoinformàtica. A: Humberto Hernández (coord.), *Aspectos de lexicografia contemporània* (p. 121-137). Barcelona: Biblograf.
- Ferret, Mariona. (2014). L'Optimot: els dubtes més freqüents i les fonts més consultades. *Llengua i Ús*, 55, 61-70.
- García Platero, Juan Manuel. (2004). Polisemia, homonimia y diccionarios. A: Josefina Prado Aragonés i M.^a Victoria Galloso Camacho (ed.), *Diccionario, léxico y cultura* (p. 131-143). Huelva: Universidad de Huelva.
- Gelpí, Cristina. (2003). El estado actual de la lexicografía: los nuevos diccionarios. A: Antonia M.^a Medina Guerra (coord.), *Lexicografía española* (p. 307-328). Barcelona: Ariel.
- Ginebra, Jordi. (2007). Una primera anàlisi del DIEC2. *Llengua i Ús*, 40, 36-43.
- Heid, Ulrich, i Zimmerman, Jan Timo. (2012). Usability testing as a tool for e-dictionary design: collocations as a case in point. A: Ruth Vatveldt Fjeld i Julie Matilde Torjusen (ed.), *Proceedings of the 15th EURALEX International Congress* (p. 661-671). Oslo: Universitetet i Oslo.
- Kaneta, Taku. (2011). Folded or unfolded: eye-tracking analysis of L2 learners' reference behaviour with different types of dictionary interfaces. A: Kaoru Akasu i Satoru Uchida (ed.), *Lexicography: Theoretical and Practical Perspectives. Papers submitted to the Seventh ASIALEX Biennial International Conference, Kyoto, August 22-24 2011* (p. 303-312). Kyoto: The Asian Association for Lexicography.
- Lew, Robert. (2012). How can we make electronic dictionaries more effective? A: Sylviane Granger i Magali Paquot (ed.), *Electronic Lexicography* (p. 343-362). Oxford: Oxford University Press.
- Nesi, Hilary. (2013). Researching users and uses of dictionaries. A: Howard Jackson (ed.), *The Bloomsbury Companion to Lexicography* (p. 62-74). Londres: Bloomsbury.

- Sanmarco Bande, María Teresa. (2008). La lexicografía del siglo XXI: hacia el diccionario ilimitado. A: José Ignacio Pérez Pascual, Mar Campos Souto i Eva María González González (coord.), *La lexicografía bilingüe y didáctica: ayer y hoy* (p. 133-140). La Corunya: Universidade da Coruña.
- Simonsen, Henrik Køhler. (2011). User consultation behaviour in internet dictionaries: An eyetracking study. *Hermes*, 46, 75-101.
- Tono, Yukio. (2011). Application of eye-tracking in EFL learners' dictionary look-up process research. *International Journal of Lexicography*, 24(1), 1-30.
- WICCAC (2018). Baròmetre de l'ús del català a Internet. Consultat el 25 de juny de 2018 des de <http://wiccac.cat/resum.html>