

# Tot el català, a través d'Internet

El català escrit dels segles XIX i XX, literari i no literari, que va ser reportat per l'Institut d'Estudis Catalans a través del buidatge de textos fet pel Corpus Textual Informatitzat de la Llengua Catalana, s'ha convertit en un 'Diccionari descriptiu de la llengua catalana' que ja es pot consultar per Internet. No té caràcter normatiu.

**E**l *Diccionari de la llengua catalana de l'Institut d'Estudis Catalans (DIEC)*, que sovint repeteix les definicions del Fabra a ulls clucs, diu que "germanor" és un "lligam entre germans" i prou. L'IEC ha vist ara que calia canviar la definició, tot i que potser Fabra, com es veu a l'article "germà", pensava que el lector inclouria en el sentit del mot germà no només els germans de sang sinó també les persones que s'hi consideraven sense ser-ho, ni que fos en situacions concretes.

La nova definició diu "unió d'afecte i solidaritat [entre persones d'un grup]" i és hostatjada des del dia 14 a la pàgina web de l'IEC. El canvi és resultat d'un projecte ambiciós, endegat el 1985.

Els lingüistes de l'IEC han vist que calia canviar la definició de "germanor" del *DIEC* i del Fabra arran del buidatge, mot a mot, nou o repetit, de cadascun dels 3.299 textos, literaris i no literaris, seleccionats de manera que fossin prou representatius de gèneres, temes i àmbits, d'entre els impresos entre els anys 1832 –la Renaixença– i 1988.

La llistata obtinguda es va tancar en arribar exactament als 52.371.994 mots, entre els quals es troben les paraules més usades durant els segles XIX i XX. Aquests mots van conformar l'anomenat Corpus Textual Informatitzat de la Llengua Catalana. La informatització del corpus va donar lloc a un *Diccionari de freqüències*, publicat, en tres volums, en paper i CD-ROM, del 1996 al 1998. És un diccionari que permet comprovar les vegades que un mot es repeteix dins dels textos seleccionats –una mostra representativa del català imprès del XIX i del XX–, és a dir, si un mot s'usa amb

més freqüència o no que un altre, i exactament en quina proporció.

Però ara la notícia, la gran notícia, és que aquest corpus textual s'ha de fer servir per crear un diccionari d'accés públic, un gran diccionari en construcció que des del dia 14 de gener es pot consultar fàcilment per Internet a través de la pàgina web de l'IEC. És un diccionari virtual que ja ha incorporat l'article "germanor". És el *Diccionari descriptiu de la llengua catalana* i és una obra viva que creix dia a dia. A data de 18 de gener, constava de 18.291 articles, mentre que a la darrera del 2004 era de 18.111.

"El diccionari tindrà finalment unes 92.000 entrades", vaticina Joaquim Rafel (Barcelona, 1943), director del projecte, membre de l'IEC i gran coneixedor de l'obra de Pompeu Fabra, sobretot de les seves *Converses filològiques*. "Si el ritme de treball continua com ara –ens diu Rafel–, trigarà uns catorze anys més a quedar enllestit."

L'equip és format a hores d'ara per 22 persones que ocupen tota una planta de les noves dependències ampliades de l'IEC, inaugurades al carrer Maria Aurèlia Capmany el 2000. L'equip inicial es va crear l'any 1998.

"Seria desitjable –continua– que el ritme de feina es mantingués i fins i tot que es fes una injecció de recursos que permetés intensificar-lo perquè l'obra es pugui enllestir en un termini que es pugui considerar raonable."

El navegant podrà accedir-hi en obrir la pàgina web de l'IEC, sota la indicació *Diccionari del català contemporani*. Aquesta denominació no bateja un diccionari concret, sinó que és un projecte lexicogràfic format pel Corpus Textual

Informatitzat de la Llengua Catalana i pel *Diccionari descriptiu de la llengua Catalana*.

El projecte també engloba una tercera obra, l'estudi, el buidatge, dels tretze diccionaris més importants apareguts de la Renaixença ençà, l'època que van ser publicats els textos del Corpus. Aquesta base de dades s'anomena BDLex, no és de consulta pública, ja que els diccionaris que l'alimenten sí que ho són, i serveix per a enriquir i afinar les definicions del *Diccionari descriptiu*, és a dir, és una eina de treball dels redactors.

“Entenem per diccionari descriptiu —ens diu Joaquim Rafel— aquella obra lexicogràfica que té per objecte la definició i la caracterització de les unitats lèxiques de la llengua des del punt de vista del contingut i de la utilització reals, sense restriccions basades en criteris prescriptius o d'una altra naturalesa. Per tant, l'obra conté mots, sentits, construccions sintàctiques, etc., que no són reconeguts per la normativa vigent, malgrat que es troben documentats en els textos. La Secció Filològica de l'IEC va decidir, després d'un ampli debat, que s'identificaria amb un punt ben visible tots els elements, sigui l'entrada, la categoria, l'accepció, el patró sintàctic etc., que no gaudeixin de la sanció normativa en el moment de divulgar-los.”

Tot i que el *Diccionari descriptiu* utilitza el Corpus Textual com a base documental i font exclusiva, cada article incorpora al final un apartat específic que recull informació lèxica obtinguda dels tretze diccionaris consultats que no s'han documentat en el Corpus.

“Així tenim reunits en una mateixa obra, però adequadament diferenciats —ens explica Rafel—, d'una banda, el resultat de l'anàlisi sistemàtica de les dades empíriques procedents del Corpus Textual, i d'una altra, tota la informació continguda en els diccionaris catalans que no té un reflex en el Corpus.”

Com que no tenen les limitacions d'espai d'un diccionari imprès, als articles hi ha una gran quantitat d'informació. N'hi ha que podria ocupar fins a quatre pàgines d'un volum. Els articles que s'ofereixen per consultar són els que es donen per acabats, al marge del seu ordre alfabètic, però molts d'aquests són sus-



Un equip de 22 persones dirigit per Joaquim Rafel s'encarrega d'anar introduint les 92.000 entrades que, segons els càlculs, aplegarà el diccionari quan estigui enllestit.

ceptibles d'experimentar certs canvis fins a l'assoliment de la seva forma definitiva.

La capçalera de l'article inclou l'entrada, la categoria, la informació flexional, la informació estadística procedent del Corpus i el perfil morfològic. El cos consta de les accepcions, les locucions, les variants, els derivats i la informació complementària. Dins del cos de l'article, el bloc d'accepcions té una certa complexitat. La descripció d'una accepció consta de diversos elements, del patró sintàctic i la definició, dels exemples i de les conversions sintàctiques.

La resposta dels primers dies sembla justificar la necessitat de l'obra. Ben aviat s'hi van registrar centenars i cente-

nars de consultes de mots per dotzenes i dotzenes d'usuaris, als quals s'adjudica automàticament, i per correu electrònic, un nom i una contrasenya.

Des del punt de vista de la seva naturalesa i del seu contingut, el *Diccionari descriptiu de la llengua catalana* s'aparta en molts aspectes dels diccionaris corrents, perquè no té un caràcter pragmàtic ni una intenció fonamentalment pedagògica. “Sense voler ser un diccionari teòric —ens diu Joaquim Rafel— comparteix amb aquests tipus d'obres algunes característiques, perquè és concebut com una investigació sobre unes bases noves, busca el màxim rigor científic, aspira a un elevat grau d'explicitud i utilitza una certa formalització en el llenguatge i en la presentació.”

Amb l'elaboració d'aquest diccionari, l'IEC, a més de produir una obra concebuda d'acord amb els principis més correntment seguits per la lexicografia moderna, es dota d'un instrument molt valuós a l'hora d'exercir la seva activitat com a institució acadèmica encarregada de l'establiment i de l'actualització de la normativa lingüística. Aquest instrument hauria també d'ajudar a elaborar finalment un diccionari normatiu rigorós que prestigii l'entitat i els filòlegs que en són membres.

**Amb aquest  
diccionari, l'IEC  
es dota d'un  
instrument molt  
valuós per exercir  
la seva activitat  
acadèmica**

Lluís Bonada