

## EL DICIONARI DESCRIPTIU DE LA LLENGUA CATALANA

Joaquim RAFEL I FONTANALS  
Universitat de Barcelona  
Institut d'Estudis Catalans

El *Diccionari descriptiu de la llengua catalana* constitueix la segona fase del projecte anomenat *Diccionari del català contemporani*. Els orígens d'aquest projecte se situen al principi de la dècada dels anys 80, en què, després d'un llarg període de forçada inactivitat, seguit d'un altre d'existència merament testimonial, l'Institut d'Estudis Catalans s'incorpora progressivament a una vida institucional ordinària i la Secció Filològica d'aleshores es preocupa per reprendre l'activitat lexicogràfica interrompuda anys abans. El desig de la Secció Filològica fou, en aquell moment, que la represa de la seva tasca creativa es produís d'acord amb els canvis que havia experimentat la societat durant un parèntesi tan llarg, i, més concretament, que els treballs que havia d'emprendre en aquest nou període de la seva història no fossin insensibles als avenços científics, metodològics i tecnològics que havien tingut lloc durant els anys de pràctica inactivitat o d'activitat precària.

El projecte *Diccionari del català contemporani* té dos objectius fonamentals que es concreten en dues fases. La primera té com a finalitat la creació de recursos lingüístics i es desenvolupà completament entre els anys 1985 i 1997; consistí fonamentalment en la constitució d'un corpus de referència de la llengua catalana escrita en els darrers cent cinquanta anys aproximadament: el *Corpus textual informatitzat de la llengua catalana (CTILC)*, l'extensió del qual supera els 52.000.000 mots; aquest corpus permet utilitzacions múltiples no solament en el camp de la lexicografia, sinó també en la recerca sobre altres aspectes de la llengua; constitueix, per tant, una infraestructura bàsica de recerca lingüística. Una realització complementària dins aquesta primera fase fou la constitució de la *Base de dades lexicogràfica*, que conté, degudament informatitzats, codificats i disposats en forma de base de dades, els tretze diccionaris catalans que han estat considerats més significatius dels segles XIX i XX; aquesta base de dades fou concebuda com un recurs auxiliar dins el projecte general i té una utilitat específica en l'elaboració del diccionari descriptiu, que veurem tot seguit. La segona fase del projecte *Diccionari del català contemporani*, que és en curs avançat de realització, consisteix en l'elaboració d'un diccionari descriptiu, el *Diccionari descriptiu de la llengua catalana (DDL)*, basat fonamentalment en l'esmentat corpus, a partir de la idea, cada vegada més estesa, que l'estudi sistemàtic dels diferents contextos en què un mot ocorre realment és la millor manera d'establir el significat de les diferents unitats lèxiques.

En una nota breu com aquesta, no podem comentar amb detall ni exemplificar cada un dels aspectes de l'estructura del *DDL*C; em referiré, per tant, només a les seves característiques generals, a l'estructura dels articles i a l'estat de la redacció. Una descripció més detallada dels diferents components de l'obra i del projecte *Diccionari del català contemporani*, de què forma part, es pot trobar a Rafel 2007.

El *DDL*C 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'altra naturalesa. L'elaboració d'un diccionari descriptiu d'aquestes característiques per part d'una acadèmia de la llengua es justifica sobretot per la creença que les decisions prescriptives seran tant més ben fonamentades com més ben coneguda sigui la llengua en la seva realitat fàctica; en el cas de l'Institut d'Estudis Catalans la justificació és doble, per tal com entre les seves missions estatutàries té no solament l'establiment de la normativa lingüística, sinó també "ocupar-se de l'estudi de la llengua". Amb l'elaboració d'un diccionari d'aquestes característiques l'Institut d'Estudis Catalans no sols produeix una obra concebuda d'acord amb els principis assumits per la lexicografia contemporània més avançada, sinó que 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.

Des del punt de vista de la seva naturalesa i del seu contingut, el *DDL*C s'aparta en molts aspectes dels diccionaris corrents; sense voler ser un diccionari teòric, comparteix amb aquest tipus d'obres algunes característiques, perquè és concebut com a una investigació sobre unes bases noves, persegueix el màxim rigor científic en la seva realització, aspira a un elevat grau d'explicitud i utilitza una certa formalització en el llenguatge i en la presentació. Com a conseqüència d'això, els destinataris ideals són les persones amb un cert domini dels conceptes lingüístics: això no obstant, no pretén adreçar-se només als especialistes, sinó que, a més de ser útil per a aquests, aspira a estar a l'abast de qualsevol usuari mitjanament culte interessat pels problemes de la llengua; pretén, doncs, combinar, en la mesura possible, el rigor en el tractament de la informació amb la claredat expositiva i amb la facilitat d'interpretació.

Una de les qüestions que planteja l'elaboració d'un diccionari descriptiu per part d'una acadèmia de la llengua és que aquesta obra conté elements (certes entrades del diccionari, determinades acepcions o sentits, alguns patrons sintàctics, certes locucions, etc.) que no són explícitament reconeguts per la normativa vigent, malgrat que es troben documentats en els textos. D'una banda, un diccionari d'aquesta naturalesa pot ser considerat més científic que un de normatiu, per tal com intenta donar raó d'una manera sistemàtica de la realitat de la llengua a partir de l'anàlisi de dades empíriques, però, de l'altra, pot ser vist com un perill per a l'ús lingüístic considerat correcte en virtut de les convencions acceptades. En el cas que ens ocupa, aquest aspecte fou àmpliament debatut en el si de la Secció Filològica, la qual decidí que s'identificaria amb una marca visible tots els elements (entrada, categoria lèxica, acepció, patró sintàctic, locució, etc.) que no gaudeixin de la sanció normativa explícita en el moment de divulgar-los.

Una de les característiques més destacades del *DDL*C, que conforma tota l'obra, és el fet d'utilitzar el *CTIL*C com a base fonamental i font exclusiva; es tracta, doncs, en sentit estricte, d'un diccionari basat en corpus. D'acord amb aquest principi metodològic, s'es-

tableix no solament la nomenclatura del diccionari (és a dir, el conjunt d'entrades que hi són tractades), sinó també els significats de les unitats lèxiques que hi corresponen, a partir de l'ús real que se'n fa; això contrasta amb el mètode utilitzat en la major part dels diccionaris, elaborats amb criteris tradicionals, que se solen basar generalment en mètodes apriorístics (el recurs als altres diccionaris existents com a font principal) o intuïtius (la consciència lingüística dels lexicògrafs). L'aplicació d'aquest mètode permet l'ús d'informació de caràcter estadístic que pot arribar a tenir una gran importància tant en la fixació de la nomenclatura com en aspectes interns de l'estructura del diccionari; la utilització sistemàtica del corpus permet la il·lustració dels diferents valors significatius a través de testimonis d'ús acreditats per una citació i una referència concretes; també permet d'incorporar al diccionari una informació exhaustiva i sistemàtica sobre les estructures sintàctiques que corresponen als mots tractats, en les seves diverses accepcions, i inventariar les principals combinacions entre els diferents mots de la llengua, d'acord amb el principi, cada vegada més sòlidament afermat en la recerca sobre el lèxic, que una de les propietats que caracteritzen les unitats lèxiques d'una llengua és la seva capacitat de combinar-se amb certes unitats del sistema amb preferència a unes altres.

Una altra característica destacada del *DDLC* es refereix al procés de redacció. A diferència del criteri adoptat en l'elaboració de la major part de diccionaris, la redacció d'aquesta obra no segueix estrictament l'ordre alfabètic, sinó que, a partir d'una progressió alfabètica bàsica, conjuntament amb l'article de l'entrada que correspon per l'ordre alfabètic, es redacten també els articles relatius a les entrades que hi estan vinculades per la seva forma o per una relació de caràcter semàntic, sigui quina sigui la lletra de l'alfabet a què pertanyen; amb això s'intenta d'evitar en bona part la manca de coherència estructural i descriptiva que afecta greument la immensa majoria de les obres lexicogràfiques existents. Així, per exemple, quan per progressió alfabètica ha tocat redactar el mot *adulació*, s'ha tractat simultàniament, per part del mateix redactor, un conjunt de 46 lemes seleccionats d'acord amb aquests criteris, és a dir, els mots relacionats formalment amb aquest (*adulador*, *adulaire*, *adular*, *adulatori*) i altres que comparteixen amb ells almenys alguna accepció (com *afalagar*, *alabar*, *elogiar*, *lloar*), i també els relacionats formalment amb aquests darrers (entre els quals, *afalac*, *afalagador*, *afalagadorament*, *afalagament*; *alabança*, *alabanciós*, *alabador*; *elogi*, *elogiable*, *elogiador*, *elogiós*, *elogiosament*; *lloa*, *lloable*, *lloablement*, *lloador*, *lloança*).

El *DDLC* conté, per a cada entrada, informació de caràcter quantitatiu relativa a l'ús observat en el corpus; aquesta informació apareix simplificada, amb una representació gràfica, no numèrica, que indica la pertinença de l'entrada a un dels cinc rangs d'ús establerts. Per a les entrades corresponents als tres primers rangs, el diccionari inclou, a més, informació sobre la repartició percentual de les categories morfològiques en el corpus (el nombre en els substantius, i la persona, el nombre, el temps i el mode en els verbs).

El *DDLC* està constituït per una sèrie d'elements estructurals organitzats en forma de base de dades, cada un dels quals està vinculat a una unitat de referència, que anomenem entrada; això permet de configurar-los en forma d'article de diccionari, que és la manera més habitual de presentar-los, per bé que no l'única possible; l'article és, doncs, la unitat bàsica que aglutina tota la informació del diccionari. Els criteris per a la delimitació dels

articles es basen exclusivament en principis de caràcter formal i gramatical i s'apliquen d'una manera sistemàtica al llarg de tota l'obra sense cap excepció.

L'article així concebut consta d'una sèrie d'elements que en constitueixen la capçalera (l'*entrada*, la *categoria*, la *informació flexional*, la *informació estadística* i el *perfil morfològic* quan escau) i d'uns altres que en constitueixen el cos (les *accepcions*, les *locucions*, les *variants*, els *derivats* i la *informació complementària*).

Dins el cos de l'article, el bloc d'accepcions té una certa complexitat; la descripció d'una accepció consta de diversos elements: el *patró sintàctic*, les *restriccions semàntiques dels elements del patró*, la *definició*, les *col·locacions*, l'*exemple* o els *exemples* i les *conversions sintàctiques*. Sense entrar en detalls ni en disquisicions tècniques, només diré que l'especificació dels patrons sintàctics ens informa sobre les diferents estructures en què ocorren les unitats lèxiques tractades, les restriccions semàntiques aporten una informació cabdal sobre les possibilitats combinatòries de les unitats dels sistema lèxic, la definició fa explícit el significat de la unitat lèxica mitjançant una paràfrasi de l'entrada, les col·locacions aporten informació sobre les combinacions més habituals del mot tractat amb altres mots, especialment d'aquelles que són el resultat d'algun tipus d'atracció lèxica; els exemples tenen un valor especial en aquesta obra: són els testimonis d'ús que il·lustren sense excepció, amb citacions extretes del *CTILC*, tots els sentits i construccions sintàctiques consignats en el diccionari; les conversions sintàctiques testimonien els usos no lexicalitzats dels infinitius verbals amb valor de substantiu i dels participis amb valor d'adjectiu.

Pel que fa a les accepcions, només afegiré que, d'acord amb la naturalesa d'aquesta obra, les definicions miren d'evitar la informació de caràcter enciclopèdic (basada en la descripció de la realitat) i es concentren en la informació de caràcter lingüístic (el valor significatiu, les restriccions semàntiques i les propietats sintàctiques de la unitat lèxica). Des del punt de vista de la formulació de la definició, es té en compte la distinció entre els elements que reflecteixen el significat conceptual en el sentit més estricte (constituents intrínsecs de la definició) i els que es refereixen a condicions o restriccions selectives (constituents extrínsecs), que reflecteixen dins la definició l'estructura argumental de la unitat lèxica definida. A Rafel 2006 es pot trobar una explanació d'aquests conceptes.

L'article inclou un apartat, titulat *Locucions* on es tracten les unitats lèxiques constituïdes per més d'un mot d'una manera anàloga a com es fa amb l'entrada. Dins aquest apartat trobem locucions nominals, com *esperit de vi*, *aigua règia* o *ull de bou*, locucions verbals, com *fer temps*, *donar a llum* o *aixafar la guitarra*, locucions adjectivals, com *dur de pelar* o *de pa sucat amb oli*, locucions adverbials, com *a dojo* o *de matí*, locucions preposicionals, com *per mitjà de* o *a tocar de*, locucions conjuntives, com *per tal com*, *sempre que* o *tan bon punt*, o locucions interjectives, com *vatua l'olla*, *ai las* o *ara pla*.

En un apartat propi de l'article es recullen les diferents variants de l'entrada documentades en el corpus; distingim entre les variants merament formals, els derivats apreciatius (diminutius, augmentatius, intensius) i les variants de flexió que no es corresponen amb la flexió normativa.

Una part important de l'article és la dedicada a la derivació. En aquest apartat s'inclouen tots aquells derivats formats segons un procés sistemàtic i que, per tant, tenen un significat previsible, a partir de la base lèxica sobre la qual es formen i el procés derivatiu de

què es tracta en cada cas. Per a cada derivat s'indica el procés derivatiu corresponent i s'addueixen tants exemples com acepcions de l'entrada són actualitzades en el corpus per aquest derivat, indicant en cada cas el nombre de l'accepció de l'entrada que actualitza. Aquest apartat inclou també col·locacions i conversions sintàctiques del derivat, si és el cas.

Finalment, l'article incorpora en un apartat específic, anomenat *Informació complementària* tota la informació lèxica relacionada amb l'entrada que es troba en algun dels diccionaris que constitueixen la *Base de Dades Lexicogràfica*, però que no s'ha documentat en el corpus; en cada cas s'indica el diccionari de procedència. Així tenim reunits en una mateixa obra, però adequadament diferenciats, d'una banda el resultat de l'anàlisi sistemàtica de les dades empíriques procedents del corpus textual, que constitueixen el cos fonamental de l'article, i de l'altra tota la informació continguda en aquests diccionaris catalans que no té un reflex en el corpus.

Des del punt de vista de la seva estructura, el *DDLC* es pot definir, doncs, d'una manera simplificada, com una base de dades multifuncional que conté per a cada entrada una sèrie d'elements autònoms relacionats de manera jeràrquica; aquests elements es corresponen d'una manera rigorosa amb els diferents components estructurals de l'article. Aquesta estructura de dades permet utilitzacions diverses, tant relacionades amb la recerca sobre la llengua com amb la difusió de l'obra; si ens centrem en la difusió, permet les més diverses possibilitats: bàsicament, una obra impresa de manera convencional, una edició electrònica independent (publicada en un disc compacte, per exemple) o bé la consulta per Internet; qualsevol d'aquestes sortides pot contenir, d'altra banda, tota la informació associada a un article o només una part, i, a més, la informació pot ser disposada físicament de manera diversa.

La idea de divulgar durant la realització del diccionari la part ja redactada de l'obra era latent des del seu inici, per tal com l'estructura de dades que la constitueix ho permet sense dificultat. Aquesta idea va acabar de prendre cos després de valorar positivament la importància d'un procediment de difusió com aquest, que s'avé amb l'esperit general del projecte; la publicació electrònica dinàmica i l'accés telemàtic a les dades del *DDLC* no solament s'integren d'una manera adequada al pla de treball de l'obra, sinó que s'ajusten plenament al plantejament innovador que el projecte té des del seu origen.

L'edició electrònica del diccionari fou posada a disposició del públic en el web de l'Institut el mes de gener de 2005, quan el nombre d'articles redactats era de 18.000; avui (març de 2014) el nombre d'articles redactats i consultables en l'edició electrònica és de 90.000, que corresponen a més del 90% de l'extensió prevista de l'obra. Aquesta edició electrònica consta d'un component nuclear, constituït pel conjunt d'articles redactats i la informació que duen associada, i uns components accessoris, que són una *Presentació*, molt breu i sintètica, i una *Guia d'utilització*, més llarga i complexa, que orienta l'usuari i li forneix les claus interpretatives.

L'accés a la informació del component nuclear es fa sempre a través de l'element que constitueix l'entrada d'un article; així, doncs, el sistema de consulta no s'allunya del d'un diccionari convencional, i és, per tant, familiar al possible usuari; però la flexibilitat que ofereix l'edició electrònica ha permès d'incorporar-hi certs vincles actius que faciliten la relació entre elements d'articles diferents i l'accés a complements d'informació que, si fossin ubicats a l'interior de l'article, n'afeixugarien el contingut.

Com que les consultes es duen a terme directament sobre la base de dades del diccionari, l'edició electrònica està permanentment actualitzada: d'una banda el nombre d'articles consultables augmenta seguint el ritme de redacció de l'obra, i de l'altra s'hi incorpora automàticament qualsevol modificació que es faci sobre la part ja redactada. Si considerem el conjunt d'articles redactats com un text en format convencional de diccionari imprès, el text corresponent tindria una extensió d'unes 70.000.000 posicions tipogràfiques. Com a punt de referència o de comparació, podem dir que, publicat en forma impresa convencional, aquest text correspondria a sis o set volums com els de l'actual *Diccionari de la llengua catalana* de l'Institut (DIEC).

Pel que fa a la distribució alfabètica, aproximadament la meitat d'aquest volum correspon als articles redactats segons la progressió alfabètica rigorosa, que arriben fins a final de la lletra *P*, mentre que l'altra meitat correspon a entrades de les altres lletres de l'alfabet associades a les primeres per raó de les similituds a què he fet referència més amunt.

La consulta del diccionari es pot fer dins el web de l'Institut d'Estudis Catalans (<http://www.iec.cat>) o bé directament a l'adreça <http://dcc.iec.cat/ddlc>.

## REFERÈNCIES

- RAFEL, Joaquim (2006): «Els elements extrínsecs en les definicions lexicogràfiques: teoria i aplicació» in BERNAL, Elisenda & DECESARIS, Janet (ed.): *Palabra por palabra. Estudios ofrecidos a Paz Battaner*. Barcelona: Universitat Pompeu Fabra; Institut Universitari de Lingüística Aplicada, 2006, p. 201-218.
- RAFEL, Joaquim (2007): «Prescripción y descripción en la actividad académica: el *Diccionario descriptiu de la llengua catalana*», in CAMPOS, Mar & al. (ed): *Reflexiones sobre el diccionario* [Actes del I Congreso Internacional de Lexicografía Hispánica, A Coruña, setembre de 2004], *Revista de Lexicografía*, Anexos 6. A Coruña: Universidade da Coruña, 2007, p. 9-33.