

ANY MOSSÈN ALCOVER
PRESENTACIÓ DE LA
LLETRA DE CONVIT 2001
LA INFORMATITZACIÓ
DEL *DICCIONARI CATALÀ-VALENCIÀ-BALEAR,*
PER
JOAQUIM RAFEL,
MEMBRE
DE LA SECCIÓ FILOLÒGICA
DE L'INSTITUT D'ESTUDIS CATALANS

Com a resultat d'una feliç coincidència de circumstàncies, l'Institut d'Estudis Catalans ha pogut iniciar el proppassat mes de desembre la tasca d'informatització del *Diccionari català-valencià-balear* d'Antoni M. Alcover i Francesc de B. Moll.

L'Institut d'Estudis Catalans cregué fa ja uns anys que per a poder desenvolupar la tasca d'estudi de la llengua i de fixació de la normativa lingüística d'acord amb les exigències dels mètodes científics moderns havia de prendre sota la seva responsabilitat la creació d'una sèrie de recursos lingüístics, necessaris per a tenir un coneixement prou complet i contrastat de certs aspectes de la realitat de la llengua, a fi de poder exercir de la manera més adequada la seva funció.

Dins aquesta línia es creà el *Corpus textual informatitzat de la llengua catalana* (entre el 1985 i el 1997), que conté més de tres mil obres amb un conjunt de més de cinquanta milions de mots, i es creà també (entre el 1995 i el 1997) una base de dades lexicogràfica que conté degudament informatitzats i carregats en una base de dades onze diccionaris catalans dels segles XIX i XX. No oblidem que la informatització de diccionaris publicats inicialment en forma de llibre constitueix una de les actuacions més destacables dins les línies generals de producció dels anomenats *recursos lingüístics*.

En el moment de crear aquesta base de dades lexicogràfica l'Institut hi dedicà els recursos de què en aquell moment podia disposar, que, tot i que eren relativament importants, eren limitats; així, incloguérem en aquesta base de dades una sèrie de diccionaris, però d'altres hagueren de quedar-se fora del projecte inicial. Entre aquests darrers hi ha, per exemple, el de Martí Gadea, el d'Aladern, i sobretot, el *Diccionari català-valencià-balear*, que, malgrat la seva importància dins la lexicografia catalana, hagué de restar a l'espera que les circumstàncies permetessin d'escometre el projecte.

Per *informatització* d'un diccionari entenem no solament el trasllat a un suport magnètic de la informació escrita que conté —aquesta és un primera fase, que podem anomenar pròpiament *digitalització*—, sinó també, i sobretot, una sèrie d'operacions entre les quals destaca per la seva importància decisiva la codificació lògica dels diferents elements estructurals del diccionari a fi de poder-lo incorporar a una base de dades i obrir-lo així a les possibilitats de tractament de la informació que la tecnologia actual ens permet. Aquestes possibilitats es concreten en aplicacions molt diverses, entre les quals podem esmentar la simple difusió i consulta per via electrònica (Internet), les accions en el camp editorial (noves plataformes de publicació, reedicions, revisions, edicions abreujades, etc.), la utilització en el camp de la lexicografia com a material complementari de consulta en projectes de nous diccionaris, l'extracció d'informació selectiva en aplicacions dins l'àmbit de l'enginyeria lingüística que requereixen la concurrència d'aquest tipus de dades, etc.

Com ja he donat a entendre, la informatització del *Diccionari català-valencià-balear* és un projecte d'una envergadura especial, no solament per l'extensió de l'obra (que s'acosta a les deu mil pàgines impreses), sinó per la complexitat de la informació que conté, tant des del punt de vista de l'estructura del seu contingut —verament complexa—, com des del punt de vista de la seva forma (pensem en els caràcters propis de la notació fonètica, o en els signes d'alfabets diferents del llatí que trobem en l'apartat corresponent a les etimologies dels mots), a part de la problemàtica específica de les taules o quadres morfològics (flexió verbal dialectal), o la munió d'il·lustracions que conté. D'altra banda, les característiques físiques del punt de partida —és a dir, de l'obra impresa— són bastant lluny de l'optimitat, cosa que no facilita la tasca de la mera digitalització; és probablement a

tot això que hem d'atribuir que algun altre intent anterior no hagi arribat a bon port.

En una intervenció com aquesta, que pretén només de donar notícia de l'execució d'aquest projecte i, en tot cas, de ponderar-ne la importància i sobretot la utilitat, no he d'entrar en detalls tècnics sobre l'execució. No és ni el lloc ni el moment adequats, com diríem, seguint el tòpic. Em limitaré, doncs, a fer un ràpid esment d'algun dels aspectes més rellevants d'aquesta actuació.

Com ja he avançat, hi ha una primera fase que consisteix a traslladar sobre suport magnètic la informació que trobem en el diccionari en forma de llibre, única versió de què disposem fins ara, amb una codificació equiparable a la tipogràfica, que, per entendre'ns, permetria reproduir l'obra tal com està impresa.

Aquest primer pas va seguit d'una operació que per una banda podríem associar a allò que en el món industrial anomenaríem *control de qualitat*, i per una altra banda inclou una sèrie d'aspectes relacionats amb la conversió de formats i de codis, a fi de disposar d'un text realment utilitzable.

Una part molt important d'aquest procés és la identificació sistemàtica dels diversos elements que configuren l'estructura dels articles del diccionari, de tal manera que puguin ser incorporats als camps corresponents d'una base de dades i, com a conseqüència d'això, puguin ser objecte de consultes orientades selectivament i puguin ser relacionats entre si.

El *Diccionari català-valencià-balear* té una complexitat estructural superior a la de la major part de diccionaris. Sense ànim de ser exhaustiu, només diré que els seus articles consten d'una entrada, que pot ser complexa; d'una informació de caràcter morfològic, que no sempre és tractable d'una manera senzilla a causa també de la seva complexitat; d'una o de diferents accepcions, fins a un nombre indeterminat,

estructurades quan convé en jerarquies de dos nivells; cada una de les accepcions consta d'una definició (que moltes vegades és múltiple, cosa que en dificulta el tractament); un o més testimonis de la llengua oral o de la llengua escrita; una referència geogràfica opcional per a la llengua oral, i una referència textual obligatòria per a la llengua escrita. Les referències geogràfiques, a part d'estar associades al testimoni oral poden referir-se a l'entrada o a una accepció i poden ser simples o múltiples (per exemple, a l'article **sorollar**, l'accepció 1, que és definida com «remenar, moure d'un costat a l'altre», està associada a les següents referències geogràfiques: el Lluçanès, la Llitera, el Priorat, Tortosa, el País Valencià). Encara dins l'àmbit de l'accepció tenim opcionalment l'equivalència o les equivalències en castellà. Fora ja de les accepcions, l'article conté un apartat opcional dedicat a la transcripció fonètica de l'entrada, que també és complex, perquè habitualment comporta més d'una transcripció, segons diferents àrees del domini lingüístic, amb les indicacions del lloc corresponent, de vegades genèriques (com valencià, balear), altres vegades específiques (Granollers, Oliana, etc.). Finalment hi ha un apartat dedicat a l'etimologia, que pot tenir una certa complexitat de cara a la codificació si es vol retenir algun dels seus elements com una informació tractable selectivament.

Un cop establerts amb claredat i d'una manera unívoca els diferents elements estructurals del diccionari que admetran un tractament individualitzat, són objecte d'un marcatge textual seguint un estàndard internacional, que permet després incorporar selectivament la informació a una base de dades i reutilitzar el text per a finalitats diverses.

Sense voler entrar en massa detalls, em referiré només de passada al fet que el conjunt d'operacions que es duen a terme durant aquest procés d'informatització inclou l'establiment de vincles lògics entre entrades que es poden considerar

variants (generalment gràfiques) associables entre si, però que no apareixen amb cap vincle associatiu en el diccionari editat. Per exemple, en la versió actual del diccionari, si consultem l'article **alegre**, no ens assabentem que el mateix diccionari inclou un article amb l'entrada **alecr**, amb el mateix significat que 'alegre', amb un testimoni d'ús procedent d'un text de l'Alguer; la versió informatitzada del diccionari permetrà obtenir directament aquesta informació perquè l'article **alecr** estarà vinculat lògicament a l'article **alegre**.

Amb això que acabo de dir, els qui m'escolteu podeu haver ja entrevist que, com a conseqüència de la informatització, accedeixen al nivell real tota una sèrie de relacions que en la versió impresa es troben en situació latent, en un nivell merament potencial. En efecte, a part de la informació que el *Diccionari català-valencià-balear* ens dona d'una manera explícita en el lloc corresponent d'un article, sota una entrada determinada, conté un gran cabal d'informació que no es troba recollida en el lloc que permetria localitzar-la segons les expectatives de l'usuari del diccionari.

Deixeu-me posar només dos exemples dels milers que podríem adduir si haguéssim finalitzat aquesta tasca d'informatització de l'obra.

A l'article **cald** (escrit antigament **calt**) -a trobem el significat d'aquest mot com a adjectiu ('calent') i com a substantiu ('calor') i uns testimonis convencionals. A través d'una consulta a la versió informatitzada, podríem investigar si en algun altre lloc no previsible del diccionari hi ha alguna informació que ens permeti completar la que es dona sobre **cald** en el seu propi article. Si cerco per *calt* (forma gràfica antiga) dins el text, trobaré que a l'article **torró** hi ha un testimoni d'ús d'aquest mot, que és una recepta per a elaborar aquesta dolça menja, tret d'un tractat de dolceria del segle XV, que s'acaba dient: «e fet asò lençar-ho has sobre una taula que sia ben neta ab aygua, e après estendràs-ho tot e fer-n'as

tauletes *de calt en calt* tals com volràs». Com veiem, això no és un mer testimoni d'ús del mot *calt* que puguem afegir als altres que hi ha en el seu article com un exemple més, sinó que és una expressió idiosincràtica (*de calt en calt*) que convindria poder recollir i explicar en l'article propi.

Un altre exemple: a l'article **somort** hi ha una expressió, **aigua somorta**, definida com «aigua que corre lentament», atribuïda a Lleida, i sense cap exemple. Si m'imagino ara la versió informatitzada que em permet cercar aquesta expressió en qualsevol lloc del diccionari, i concretament en els exemples o testimonis d'ús dels altres articles, trobaré que a l'article **regatiu** («terreny que es rega artificialment» —Rosselló, Conflent—) hi ha un exemple de l'obra *La mare terra* (Perpinyà, 1913) de l'autor rossellonès Pau Berga, que diu «del regatiu a l'aigua somorta». Amb això he obtingut *a*) un exemple d'una expressió que en el lloc on és referenciada no té exemple, i *b*) una informació complementària: que aquella expressió que el diccionari, en l'article corresponent, localitza només a Lleida té un testimoni escrit d'un autor del Rosselló. I així podríem continuar.

Deixeu-me dir —i no em quedaria tranquil si no ho fes— que tot això és normal que passi en una obra elaborada amb procediments manuals, sense el concurs d'un mitjà com l'informàtic, que permet un control sistemàtic de la informació que es maneja.

Deixeu-me dir també, abans d'acabar, que aquest pas que fem ara d'informatitzar l'obra concebuda i iniciada per mossèn Antoni M. Alcover i duta a terme per F. de B. Moll, a part de la seva importància objectiva per a l'actualització dels repertoris lingüístics de què disposem, revalorant-los mitjançant les possibilitats d'ús i d'explotació multidireccional, té un valor simbòlic. El *Diccionari català-valencià-balear* és per als catalans i per a tots els estudiosos d'arreu un gran diccionari, però per als catalans és, a més, un símbol, com ho

era ja el cedulari que conté les fitxes manuals que el precediren —la famosa calaixera—; ens ho recorda el mateix F. de B. Moll en la introducció del diccionari, quan ens diu: «Aquesta calaixera es va fer famosa i es convertí, per als catalans entusiastes, en una mena d'urna sagrada, d'arca de l'aliança de la catalanitat.»

Crec que puc dir que contribuir al fet que aquest símbol, a més de la seva presència estàtica en la versió actual, adquireixi una forma dinàmica a través de la incorporació al món dels formats electrònics és un honor per a l'Institut d'Estudis Catalans. I el fet que coincideixi la realització del projecte amb la commemoració del centenari de la *Lletra de convit* de mossèn Alcover és un motiu de satisfacció.