

L'APORTACIÓ DE LA BASE DE DADES DE LES OFICINES LEXICogrÀFIQUES DE LA SECCIÓ FILOLÒGICA A LA LEXICOGRAFIA CATALANA

Cristina Busto [Centre d'Estudis i Desenvolupaments Informàtics de l'IEC]
Carolina Santamaria [Oficines Lexicogràfiques de l'IEC]

1. INTRODUCCIÓ

En la comunicació que presentem avui ens proposem d'abordar una qüestió directament relacionada amb el *Diccionari de la llengua catalana* de l'Institut (DIEC), aparegut l'octubre de 1995: la descripció del que ha estat l'eina fonamental per a l'elaboració del DIEC, és a dir, la descripció de la base de dades de les Oficines Lexicogràfiques, que anomenem BDOL.

En primer lloc, farem, molt breument, unes reflexions generals sobre el paper que pot tenir la informàtica en la pràctica lexicogràfica. A continuació, explicarem algunes característiques de la BDOL: com va néixer, com es va dissenyar, com s'ha implementat, les parts de què és constituïda i les dades que s'hi han introduït. Seguidament, insistirem en els avantatges que suposa per a un projecte lexicogràfic el fet de disposar d'una base de dades de les característiques de la BDOL i explicarem les grans possibilitats d'explotació que ofereix. Finalment, exposarem els plans de futur que té l'IEC per a aquesta base de dades de les Oficines Lexicogràfiques.

2. ALGUNES GENERALITATS SOBRE EL PAPER DE LA INFORMÀTICA EN LA LEXICOGRAFIA

De tothom és sabut que la informàtica té múltiples aplicacions en el camp de la lexicografia. De fet, molts projectes lexicogràfics han demostrat que les eines informàtiques poden ésser un gran ajut en l'elaboració d'un diccionari.

La utilitat de la informàtica es fa evident en totes les fases de l'elaboració d'una obra lexicogràfica. El diccionari anglès Cobuild (*Collins Cobuild English Language Dictionary*, Londres i Glasgow, Collins, 1987) és un exemple clar d'obra lexicogràfica elaborada des de l'inici amb l'ajut d'eines informàtiques.

Aquest és també el cas del DIEC: la base de dades de les Oficines Lexicogràfiques ha estat imprescindible en totes les fases d'elaboració del nou diccionari:

- 1a) En la preparació, que ha consistit, fonamentalment, en la recollida de les dades.
- 2a) En la selecció de les entrades.
- 3a) En la construcció dels articles.
- 4a) En la publicació del diccionari.

Concretament, l'eina informàtica capaç de superar les limitacions que té un diccionari convencional és una base de dades relacional. En una base de dades d'aquest tipus, la informació està estructurada en

diversos arxius —no es presenta d'una manera lineal. Tots els arxius de la base estan relacionats entre si: tots tenen un camp o altre en comú. Aquesta estructura permet explicitar les relacions que hi ha entre registres i entre camps.

Alguns dels avantatges que representa per a un lexicògraf disposar d'una base de dades relacional són els següents:

— L'estructura de la base és fàcil de manejar pels usuaris.

— La base de dades lexicogràfiques permet codificar els textos introduïts per tal de poder-los manipular fàcilment.

— Permet al lexicògraf d'accedir a un volum considerable d'informacions simultàniament, atesa la gran capacitat d'emmagatzematge.

— Si s'hi introdueixen fonts lexicogràfiques ja difoses, la base de dades pot contenir un corpus de referència molt valuós per al lexicògraf.

— L'accés a les dades és ràpid i fàcil. La consulta és, doncs, àgil i, si cal, selectiva. Així doncs, les maneres d'obtenir les dades són diverses i múltiples.

— L'actualització de les dades també és ràpida. Es poden modificar els arxius sense que les relacions ni les dades en surtin perjudicades. La base pot créixer o disminuir sense que calgui modificar el sistema en conjunt, i permet tractar les dades i obtenir-ne resultats de manera immediata.

— Pot contenir programes que permetin realitzar diverses verificacions del contingut de les definicions. Són, doncs, programes que controlen la coherència interna i eviten la circularitat d'aquestes definicions.

— Pot descarregar el lexicògraf de les tasques més mecàniques; per tant, li permet de distribuir millor el temps dedicat al maneig de les dades i a la reflexió sobre aquestes.

— Simplifica el procés d'edició d'un diccionari en paper i en suport informàtic, atès que permet codificar molts aspectes de l'edició dels productes lexicogràfics: tipus de lletra, disposició espacial del text, símbols, etc.

De tota manera, no hem d'oblidar que hi ha tasques intel·lectuals inherents a la pràctica lexicogràfica que, de moment, no pot dur a terme cap ordinador —especialment, anàlisis que es fonamenten en la intuïció i en els coneixements dels lexicògrafs.

3. LA BASE DE DADES DE LES OFICINES LEXICOGRÀFIQUES

Des del primer moment en què la Secció Filològica encarregà a les seves Oficines Lexicogràfiques l'elaboració d'un nou diccionari normatiu, la directora d'aquestes Oficines, la doctora Teresa Cabré, va insistir en la necessitat d'emprar eines informàtiques en aquest nou projecte lexicogràfic, per tal de facilitar i agilitar les tasques que s'havien de dur a terme.

Aquest fou, doncs, l'objectiu fonamental de la creació de la BDOL: dotar el projecte lexicogràfic normatiu de l'Institut d'unes eines informàtiques suficients. No fou, però, l'únic objectiu. La doctora Cabré defensava la idea que la base de dades de les Oficines Lexicogràfiques havia de permetre la realització de tota mena d'estudis lèxics i el forniment de dades per a qualsevol projecte lexicogràfic. Així doncs, a partir del disseny fet per les Oficines, es creà, en el Centre d'Estudis i Desenvolupaments Informàtics de l'IEC, una base de dades lexicogràfiques relacional útil per a múltiples projectes i treballs:

— per a l'elaboració de la primera edició del DIEC, en primer lloc;

— per a les successives reimpressions i edicions del DIEC;

- per a l'elaboració d'altres projectes lexicogràfics;
- per a la realització de treballs d'investigació sobre el lèxic.

3.1. Parts de la base de dades lexicogràfica

Des del punt de vista dels usuaris, els camps en què s'estructura cada registre d'aquesta base de dades lexicogràfiques s'agrupen en quatre grans blocs, ben diferenciats segons el tipus d'informació que contenen:

- 1) Dades per a la *identificació de les entrades*.
- 2) Informacions procedents de les *fonts*.
- 3) Informacions per a les *propostes lèxiques*.
- 4) Dades de *gestió de les informacions*.

1) En el bloc corresponent a la identificació de les entrades (figura 1), trobem el camp per al *codi numèric* d'identificació del mot, que és assignat automàticament pel programa; el camp per a l'*entrada*, que és el mot de referència per a totes les fonts; el *camp de classificació*, que permet que les entrades de la base es puguin ordenar alfabèticament d'una manera correcta; el camp per a la *categoria gramatical*, i altres *dades complementàries*, comunes a totes les fonts.

2) En el bloc corresponent a les informacions procedents de les fonts (figura 2), trobem dos camps que contenen la informació específica procedent de cadascuna d'aquestes fonts. En el primer d'aquests camps, de sis caràcters, s'especifica el codi corresponent a la font de procedència de les informacions. En l'altre hi ha totes les informacions contingudes en aquesta font.

Les Oficines Lexicogràfiques han introduït a la BDOL diversos tipus de fonts. A banda de les fonts de caràcter normatiu (el *Diccionari general de la llengua catalana*, i els volums I i II dels *Documents de la Secció Filològica*), que evidentment havien de ser les obres de partida, també hi tenim vuit fonts lexicogràfiques, setze fonts de neologismes (si hi sumem també les fonts en què es recullen les propostes dels mitjans de comunicació), vuit fonts d'argot, set fonts de dialectalismes i setanta fonts terminològiques.

Les informacions corresponents a totes aquestes fonts han estat introduïdes linealment; hi ha, però, algunes dades que han estat codificades per tal que es puguin recuperar per mitjà de llistats.

3) El tercer bloc de camps que podem distingir en cada registre —el més complex— (figura 3) conté les propostes lèxiques, i reproduïx l'estructura dels articles del *Diccionari general* de Pompeu Fabra. En aquesta part de la BDOL és on les Oficines Lexicogràfiques recullen, distribuïdes en accepcions, les informacions que proposen d'incloure en l'article del futur diccionari.

En general, l'estructura de les bases de dades lexicogràfiques depèn, bàsicament, de l'estructura de les obres lexicogràfiques que s'hi introdueixen. En el cas de la BDOL, l'obra lexicogràfica que en determinà l'estructura fou el *Diccionari general de la llengua catalana*, de Pompeu Fabra (DGLC). Les Oficines van dissenyar l'estructura de la fitxa lexicogràfica de la base a partir de la manera com estaven estructurats els articles del *Diccionari general* i d'una llista exhaustiva de tots els tipus d'informacions que podien aparèixer en un article d'aquest diccionari. Val a dir que el punt de partida d'aquest disseny fou un estudi previ elaborat pel doctor Joaquim Rafel i per la doctora Victòria Alsina sobre l'estructura dels articles del DGLC.

Els camps es van disposar en una estructura que reflecteix les relacions de dependència que hi ha entre les diverses informacions dels articles del diccionari. Dins de cada accepció tenim dos nivells d'informació. En primer lloc, tenim les dades generals de l'accepció: codi de fase (indica si els continguts

de l'accepció són definitius, provisionals, si queden pendants d'estudi, etc.), data i responsable de l'atorgament d'aquesta fase, signe separador que precedeix l'accepció, categoria gramatical, àrea temàtica, valoració lingüística i informació morfològica). En un segon nivell, trobem el camp en què hi ha el text de la definició, de l'exemple o de la subentrada.

La primera font que es va introduir a la proposta de la BDOL fou, lògicament, el DGLC, atès que era el diccionari objecte d'actualització. La introducció del *Diccionari general* de Pompeu Fabra es féu de manera automàtica. Les fases d'aquesta introducció varen ser les següents:

— Escanejament del text de la segona edició del DGLC (publicada l'any 1954 i última revisada per Pompeu Fabra) al Servei de Textos Informatitzats en Llengua Catalana, de la Universitat de Barcelona. Després de l'escanejament, hom ja disposava del DGLC en suport informàtic, tot i que amb una estructura molt senzilla, lineal.

— Introducció de codis al davant de cada informació, per indicar-ne el tipus. Vegem-ne tot seguit una mostra.

[EN]	precedia sempre la forma de l'entrada i, per tant, l'inici de cada article.
[CG]	precedia sempre la categoria gramatical de l'entrada.
[IM]	precedia sempre la informació morfològica de l'accepció.
[DF]	precedia sempre el text de les definicions.
[EX]	precedia sempre el text dels exemples.
[MF]	precedia sempre les subentrades.
[FI]	indicava sempre el final dels articles.

— Elaboració d'un programa de bolcatge o *mapping* que permetés que cada informació es traspassés automàticament, gràcies als codis introduïts, al camp de la BDOL que li havia estat assignat.

— Bolcatge del DGLC a la BDOL. El traspàs es féu des d'una cinta magnètica que contenia el text escanejat del DGLC. Un cop fet el bolcatge, es féu una revisió del resultat d'aquesta operació. D'aquesta manera, el DGLC passava a tenir un format de diccionari informatitzat amb una estructura de base de dades lexicogràfiques.

4) La darrera part de la fitxa lexicogràfica de cada registre de la BDOL és la que anomenem «de gestió de les dades» (figura 4). Conté la data en què es donà d'alta l'entrada a la base, les inicials de la persona que la donà d'alta, la fase en què es troba, la data en què se li atorgà aquesta fase i les inicials de la persona que féu l'atorgament, etc.

3.2. Manteniment i consulta de la base de dades

Pel que fa al manteniment i a la consulta, cal dir que la BDOL és una base de dades en evolució constant: el manteniment no s'ha aturat en cap moment, atès que són tasques permanents de les Oficines Lexicogràfiques la introducció de noves dades i la revisió de les dades que ja han estat introduïdes.

D'altra banda, la recuperació de les dades de la BDOL pot tenir dues formes: en línia, a través dels terminals de l'ordinador IBM AS/400, o bé en suport imprès, en forma de llistats. Hi ha un gran ventall de modalitats d'impressió i de visualització de les dades.

Pel que fa a la impressió de les dades, en el menú de llistats de la BDOL trobem, entre altres, aquestes opcions:

- Llistats de mots per fonts (selecció de les entrades de la BDOL que tenen unes fonts determinades).
- Llistats de mots per categoria gramatical, per àrea temàtica, per valoració lingüística, per fases.
- Llistats de mots amb una seqüència determinada en el text de les accepcions.
- Llistats de coherència en què apareixen els mots que no són entrades del diccionari, malgrat que sí que hi apareixen en alguna definició o en algun exemple.

Per pantalla, els camins de consulta són dos:

a) Per la forma de l'entrada. Un cop teclejada l'entrada que l'usuari vol consultar, es pot accedir a totes les dades que té associades. Una de les possibilitats és la consulta de les informacions introduïdes a la proposta (figura 5), disposades com en un article de diccionari i amb una opció d'hipertext que permet obrir, des d'una entrada, fins a cinc entrades successivament, les informacions de les quals se superposen en la pantalla. També hi ha la possibilitat de consultar la fitxa tècnica de cada entrada, que permet accedir a totes les informacions de les fonts i de la proposta alhora (figura 6).

b) Pel programa de consultes. Aquest programa permet interrogar la base de dades de manera exhaustiva: es poden fer recerques en tots els camps codificats i en els camps de text lliure de la proposta, un per un o bé combinadament (figura 7).

La manera com està dissenyat aquest programa fa que els usuaris obtinguin el màxim rendiment en el mínim temps de resposta.

Per tal de fer recerques sobre els camps lliures, ha calgut seguir tot un procés d'indexació del text contingut en aquests camps. En tots els camps lliures les informacions han estat indexades mot a mot i segment a segment. En aquests camps, per tant, es poden fer recerques truncant els mots pel començament o pel final.

Només, però, en els camps per a l'entrada, per al mot alternatiu, per a la flexió de la forma del femení, per a les subentrades i per a la forma abreujada de les subentrades es poden recercar formes truncades pel començament i pel final alhora i seqüències de més d'un mot (amb espais).

Les informacions dels camps per al mot en la llengua d'origen, la pronunciació i la informació morfològica de l'entrada i de l'accepció no han estat indexades.

La interrogació de la BDOL des del programa de consultes es fa mitjançant frases booleans.

Cada frase consta d'un indicatiu, d'un signe i d'un argument. Aquestes frases es poden combinar, i donen lloc a expressions i a multiexpressions. Els operadors que permeten aquestes combinacions són *i*, *n* ('no') i *o*. En les recerques amb truncacions es fan servir, bàsicament, els signes *, que representa un caràcter qualsevol; \$, que representa sempre un espai (inici i final de mot), i =, que representa un caràcter o un espai.

Un programa verificador de la sintaxi d'aquestes frases comprova en cada consulta que les preguntes estiguin ben formulades.

El programa de consultes permet, a més, que l'usuari dissenyi la manera com vol visualitzar i imprimir el resultat de la consulta que ha fet: és a dir, que pot triar quines informacions vol veure o imprimir, en quin ordre, amb quina tipografia, etc.

El programa de consultes permet llistar sobre paper totes les consultes que s'hi fan. D'entrada, ofereix tres tipus de llistat estàndard: impressió en forma de fitxa tècnica, impressió en forma de llista i impressió en format d'article de diccionari.

3.3. Característiques i estructura interna de la base de dades

Vegem ara, des del punt de vista informàtic, quines són les característiques i l'estructura interna de la BDOL (figura 8).

La base de dades de les Oficines Lexicogràfiques està implementada en un ordinador IBM AS/400, model E70, que utilitza el sistema operatiu OS/400, versió 2, revisió 3.

El llenguatge de programació amb què s'ha construït aquesta base és l'RPG 400.

Actualment, la BDOL conté, aproximadament, 104.000 unitats lèxiques i 5 milions de registres. Ocupa, físicament, 655 megabytes de la memòria de disc dur de l'ordinador. En total, la BDOL és constituïda per 313 programes, si sumem els programes de manteniment, els de consultes i els de generació d'informes o llistats.

El nombre total d'arxius físics (que contenen dades) és de 84. A més dels arxius físics, la BDOL consta també de 40 arxius lògics (contenen les diferents rutes d'accés a les dades).

En la figura 9 podem veure, en un esquema en forma d'arbre, la jerarquia que s'estableix entre alguns dels arxius que acabem d'enumerar.

Com hem dit més amunt, la BDOL és una base de dades relacional; per tant, tots els arxius de la BDOL estan relacionats entre si: tots tenen en comú un camp amb el codi numèric que el programa assigna automàticament a les unitats lèxiques, a mesura que es donen d'alta (figura 10).

4. L'EXPLOTACIÓ DE LA BDOL

Diem, en començar aquesta exposició, que la base de dades de les Oficines Lexicogràfiques ha estat l'eina fonamental en totes les fases d'elaboració del DIEC. Vegem, ara, més detalladament, de quina manera aquesta base ha facilitat les tasques dutes a terme amb vista a la publicació del nou diccionari normatiu:

1) Les Oficines Lexicogràfiques han introduït a la BDOL les dades de les diverses fonts que s'han buidat per a l'actualització del diccionari de Pompeu Fabra. En aquest sentit, doncs, la base de dades ha estat molt útil en la primera fase d'elaboració del DIEC —la preparació i la recollida de les dades. El suport informàtic ha permès de manipular els textos d'aquestes fonts i codificar-ne algunes informacions per tal de poder-les recuperar d'una manera selectiva.

2) La introducció del DGLC a la part de les propostes lèxiques de la BDOL ha servit, gràcies a la possibilitat d'extreure'n llistats, per a detectar errors i incoherències de tota mena que hi poguésser haver en el DGLC i per a fer els estudis sobre els criteris d'elaboració del diccionari de Fabra. A partir d'aquests estudis es van poder fixar els criteris d'elaboració per al nou diccionari.

La base de dades de les Oficines Lexicogràfiques també ha permès de fer explícites informacions que eren del tot implícites en el DGLC, com ara les àrees temàtiques a què pertanyen les accepcions.

3) Pel que fa a la fase de selecció, la introducció de codis d'especialitat a cadascuna de les accepcions del DGLC ha permès d'extreure llistats d'entrades d'aquest diccionari per àrees temàtiques. També s'han extret de la BDOL llistats per àrees temàtiques d'altres fonts lexicogràfiques introduïdes, llistats de les fonts d'argot, de les fonts de dialectalismes, de les fonts en què es recullen les propostes dels mitjans de comunicació, etc. Fou a partir de tots aquests llistats que els membres de la Secció Filològica i els diversos especialistes contactats per les Oficines Lexicogràfiques van fer la selecció dels mots i de les accepcions que havien de tenir cabuda en el DIEC.

Un cop feta aquesta selecció, s'elaborà un programa informàtic que féu possible que la introducció

del resultat de la selecció fos semiautomàtica i, per tant, àgil. Atès que les Oficines treballaven aleshores sota la pressió dels terminis d'edició, és evident que disposar d'aquest programa suposà un gran estalvi de temps.

4) En la fase d'elaboració dels articles del diccionari, són moltes les tasques que la base de dades ha facilitat: homogeneïtzació d'entrades de característiques semblants, tant pel que fa a la forma com pel que fa al contingut semàntic, detecció d'errors i incoherències internes gràcies a un programa que verifica que tots els mots emprats en les definicions tinguin entrada en el diccionari, etc.

Per a la revisió final dels articles, en la qual les Oficines Lexicogràfiques van preveure que un bon percentatge de les esmenes consistiria a reordenar-ne els continguts, es dissenyà un programa que permetia tallar i enganxar les informacions de manera automàtica.

5) Finalment, la BDOL també ha facilitat les tasques relacionades amb la fase d'edició del DIEC. L'Institut proporcionà al fotocomponedor un arxiu en què cada informació anava precedida d'una etiqueta que n'indicava el tipus i la tipografia. Per tant, la fase de l'edició quedava pràcticament automatitzada.

6) Actualment, la BDOL continua sent l'eina de treball fonamental per a les tasques que duen a terme les Oficines Lexicogràfiques, la major part de les quals tenen a veure amb les reimpressions del DIEC i amb la segona edició del diccionari. Recordem que la primera i la segona reimpressions van aparèixer el març de 1996 i el gener de 1997, respectivament, i que, en aquests moments, les Oficines Lexicogràfiques ja estan duent a terme diverses tasques amb vista a la segona edició del diccionari. També hem de dir, en aquest punt, que no hauríem pogut elaborar les llistes aparegudes en el llibre de la Secció Filològica *Documents normatius 1962-1996 (amb les novetats del diccionari)*, aparegut enguany, sense l'ajut de la BDOL: l'acarament de diversos llistats extrets de la base féu possible la confecció de la llista de diferències gràfiques entre el DGLC i el DIEC, la llista dels articles del DGLC suprimits en el DIEC i la llista dels articles nous del nou diccionari normatiu.

7) Darrerament, hem vist que la BDOL pot esdevenir una font de dades important per a altres projectes. Pel que fa al projecte de la gramàtica de l'Institut, es proporcionà un arxiu amb totes les entrades del DIEC que duïen en alguna de les seves accepcions l'àrea temàtica de filologia. Aquest fitxer fou extret de la BDOL i recuperat en una base de dades en MS Access. Igualment, les Oficines Lexicogràfiques varen proporcionar, també en format d'MS Access, els termes que especialistes de la Secció de Ciències Biològiques van sol·licitar per endegar l'elaboració d'un diccionari de ciències ambientals.

8) No podem oblidar que una de les últimes aplicacions que ha permès la base de dades de les Oficines Lexicogràfiques és la consulta del DIEC a través de la xarxa Internet. L'accés es pot fer únicament per mitjà d'una connexió de Telnet. En la pàgina corresponent del web de l'Institut s'explica com pot accedir l'usuari a la consulta.

5. EL FUTUR

Els projectes de futur per a la BDOL presenten facetes ben diverses. Aviat, la base de dades de les Oficines Lexicogràfiques serà l'eina que facilitarà la selecció de les entrades i les accepcions que han de formar part de la nomenclatura del diccionari manual projectat recentment per la Secció Filològica.

Actualment, s'estudia la possibilitat de fer-hi recerques *full text* (és a dir, a tota la base de dades en conjunt) mitjançant el programa de recerca documental Topic, que ja ha estat adquirit per l'Institut.

El projecte més ambiciós és, però, la transformació de la BDOL en una base de dades lèxiques i, posteriorment, en un diccionari computacional. Les Oficines Lexicogràfiques de l'IEC són un dels grups de recerca del Centre de Referència en Enginyeria Lingüística (CREL). En el marc d'aquest Centre, les Oficines duen a terme, en col·laboració amb l'Institut Universitari de Lingüística Aplicada de la Universitat Pompeu Fabra (IULA), diverses activitats amb vista a aquest futur diccionari computacional. Així, en aquests moments s'estudia la possibilitat d'aplicar a la BDOL un programa analitzador morfològic —ja elaborat a l'IULA—, i s'han iniciat les tasques per a la programació d'un analitzador sintàctic.

Un requisit previ i imprescindible per dur a terme aquest projecte de convertir la BDOL en un diccionari computacional és la migració de la BDOL (tant de les dades com de les aplicacions) a sistemes oberts (compatibles), concretament, a una plataforma múltiple de la marca Unix. Aquesta migració ja ha obtingut el vistiplau de les instàncies corresponents de l'IEC i s'hi ha començat a treballar. L'ordinador en què s'instalarà la base de dades és una màquina SUN Ultra 2, amb 192 megabytes de memòria RAM i una capacitat en disc de 6,3 gigabytes. La base de dades actual es traspasarà a una base de dades de la marca Oracle i el llenguatge de programació serà JAVA.

Els objectius generals de la migració de la BDOL a un entorn obert de xarxa sobre la plataforma Unix són:

- La superació de les limitacions tecnològiques que representa la plataforma actual IBM AS/400.
- La millora de les aplicacions que la BDOL té actualment.
- La creació de noves aplicacions que amplii la potència de la base.
- La revisió de l'estructura de dades amb vista a la recerca d'una estructura flexible que permeti ampliar les possibilitats d'estudi i d'explotació de les informacions.

Més concretament, algunes de les millores que s'obtindran són, per exemple:

- L'ampliació dels tipus d'indexació possibles per accelerar les recerques.
- La visualització en pantalla dels articles del DIEC amb la tipografia que tenen en el text imprès del diccionari.
- L'enllaç entre les entrades i els documents en què se'n parla (actes de la Comissió Lexicogràfica, informes de les Oficines, etc.).
- Els hiperenllaços, és a dir, les connexions entre informacions d'entrades diferents, etc.

6. CLOENDA

Amb aquesta breu notícia sobre els canvis que el futur més immediat té reservats per a la BDOL, tanquem la nostra exposició, en la qual hem reflexionat sobre el paper que la informàtica pot tenir en les tasques lexicogràfiques, hem descrit la base de dades de les Oficines Lexicogràfiques i hem volgut donar a conèixer a tots els presents les prestacions d'aquesta base de dades i l'aportació que creiem que ha significat i pot significar per a la lexicografia catalana.

FIGURA 1
Dades d'identificació de les entrades

	Diccionari Normatiu	ESMENA DE MOT
	BDOL	Dades generals
1	→ Codi mot.....:	9.672
2	→ Entrada	blues
	Data d'introducció....:	17/07/91 Inicials usuari: CSJ Carolina Santamaria
	Fase	8 Mots editats
	Data de la fase	25/05/95 Resp. fase.....: 888 Lliurament del NDN (disq
3	→ Camp de classificació..:	blues
	Nro. ordre classific...:	
	Impressió de l'entrada:	
	Femení desenvolupat...:	
	Femení imprès.....:	
4	→ Categoria gramatical...:	M substantiu masculí
	Mot alternatiu.....:	
	Categoria gramatical...:	
	Llengua d'origen.....:	AN Anglès
	Mot original.....:	
	Pronunciació.....:	
	Informació morfològica:	
	[inv.]	
	1 Marca.....:	EST Mot estranger
	2 Marca.....:	MUS Gèneres musicals
	3 Marca.....:	PGR Problemes gràfics
	4 Marca.....:	
	5 Marca.....:	
	Situació FABRA origen.:	N Entrada nova
	Situació DIEC1 origen.:	

FIGURA 2
Informacions de les fonts

Diccionari Normatiu CONSULTES D'INFORMACIÓ DELS MOTS

Entrada: calor

Codi.: 15.053

S Codi Descripció/Informació

BDOL2 Còpia de BDOL en data 6.7.95

[EN]calor [CG]f. 010 [AT]LC Lèxic comú [AT]FIF Física en general [DF]* +TFISIC+ Manifestació de l'energia interna d'un s

CODOL Còpia de BDOL en data 10.5.93

[EN]calor [CG]f. [AT]LC Lèxic comú [AT]FIF Física en general [AT]ZOA Física en general [DF]Forma d'energia (probablement

DGLC Diccionari general de la Llengua Catal.

calor [CG]f. [DF]Forma d'energia (probablement l'energia cinètica de moviments oscil·latoris de les molècules de la matè

DRCF Diccionari català-francès (GEC)

f. *Fer calor*, faire chaud.

DRDLC Diccionari de la Llengua Catalana

f. 1.a. <TERMO> Manifestació de l'energia interna quan un sistema efectua un procés mitjançant canvis de temperatura, si

TFISIC Termes de física general

+

F3 sortida F12 anterior X selecció

FIGURA 3
Propostes lèxiques

Diccionari Normatiu

ESMENA DE MOT

Dades generals acceptió

ACCEPCIÓ: 40 Ordre: 40 Codi: 15.053

Entrada.: calor

Primera informació.: D Animació, passió, amb què hom fa una cos

Fase

Data de la fase....: 0/00/00 Resp. fase:

Tipus de divisió.....: 2 || DOBLE BARRA

Categoria gramatical..:

1 Àrea temàtica.....: LC Lèxic comú

2 Àrea temàtica.....:

3 Àrea temàtica.....:

1 Valoració ling.....: FIG figuradament

2 Valoració ling.....:

Situació respecte DIEC:

Informació morfològica:

F3 Sortida F9 Confirmació F12 Anterior

FIGURA 5
Consulta en forma de diccionari

calor f. Manifestació de l'energia interna d'un sistema que es mesura per mitjà de la temperatura. *Calor despresa en cremar-se un cos, condensar-se un vapor. Calor absorbida en fondre's un cos, en evaporar-se un líquid. Transmissió de la calor.*
calor animal Calor generada en el cos d'un animal pels processos d'oxidació que tenen lloc en el seu organisme. **calor latent** Calor que un cos absorbeix sense que augmenti la seva temperatura. | **calor específica d'un cos** Quantitat de calor que es requereix per a elevar un grau la temperatura d'1 gram d'un cos. || Sensació que experimenta el cos animal quan rep calor de l'exterior o quan augmenta la seva temperatura interior a causa d'un estat febril (oposat a fred). *Tenir, sentir, calor. Les calors de l'estiu. Corrent he agafat calor.* || *fig.* Animació, passió, amb què hom fa una cosa. *En la calor de la discussió, de la disputa.*

FIGURA 6
Consulta de la fitxa tècnica

Entrada.....: calor

Número...: 15.053 Fase...: 9 CG F OR AL Gestió: SLC 2/06/97
Marques.: SF M SD M Resp...: EPS 21/08/91
Morf.....:

Fonts...: BDOL2 CODOL DGLC DRCF DRDLC ESTAND TFISIC TMEDIC TZOOL
Total d'accepcions: 006 Número de l'última accepció: 040

Acc.....: 10 Fase...: CG AT LC FIF VAL DIV SD q
Morf.....:

DF Manifestació de l'energia interna d'un sistema que es mesura per mitjà de la temperatura.
EX Calor despresada en cremar-se un cos, condensar-se un vapor.
EX Calor absorbida en fondre's un cos, en evaporar-se un líquid.
EX Transmissió de la calor.

Ampliació: de l'acc.:

F3 sortida F12 anterior F15 + accepcions F16 - accepcions F24 forma diccionari

FIGURA 7
Programa de consultes

1/12/97

CONSULTES - DICCIONARI NORMATIU

1 sel fase=8,8r,9 i cg=f i at=fif,ei,ee

658 Mots seleccionats

2 sel fase=8,8r,9 i en=coma

4 Mots seleccionats

3 vco

4 sel fase=8,8r,9 i en=("**derma")

12 Mots seleccionats

5 vco

6 sel fase=8,8r,9 i en=("hiper**")

54 Mots seleccionats

7 sel fase=8,8r,9 i dfc=escola

17 Mots seleccionats

Indiqueu ordre: sel fase=8,8r,9 i dfc=escola

F3 fi de consultes

Ajuda

F8 recuperar

F9 confirmar

F11 pant.completa

FIGURA 8
Algunes dades sobre la BDOL

ORDINADOR:	IBM AS/400, model E70
SISTEMA OPERATIU:	OS/400, versió 2, revisió 3
LLENGUATGE DE PROGRAMACIÓ:	RPG 400
NOMBRE D'UNITATS LÈXIQUES:	104.000
NOMBRE DE REGISTRES:	5.000.000
OCUPACIÓ DE MEMÒRIA (bytes):	655.000.000
NOMBRE DE PROGRAMES:	313
NOMBRE D'ARXIU FÍSICS:	84
NOMBRE D'ARXIU LÒGICS:	40

FIGURA 9
Jerarquia dels arxius principals de la BDOL

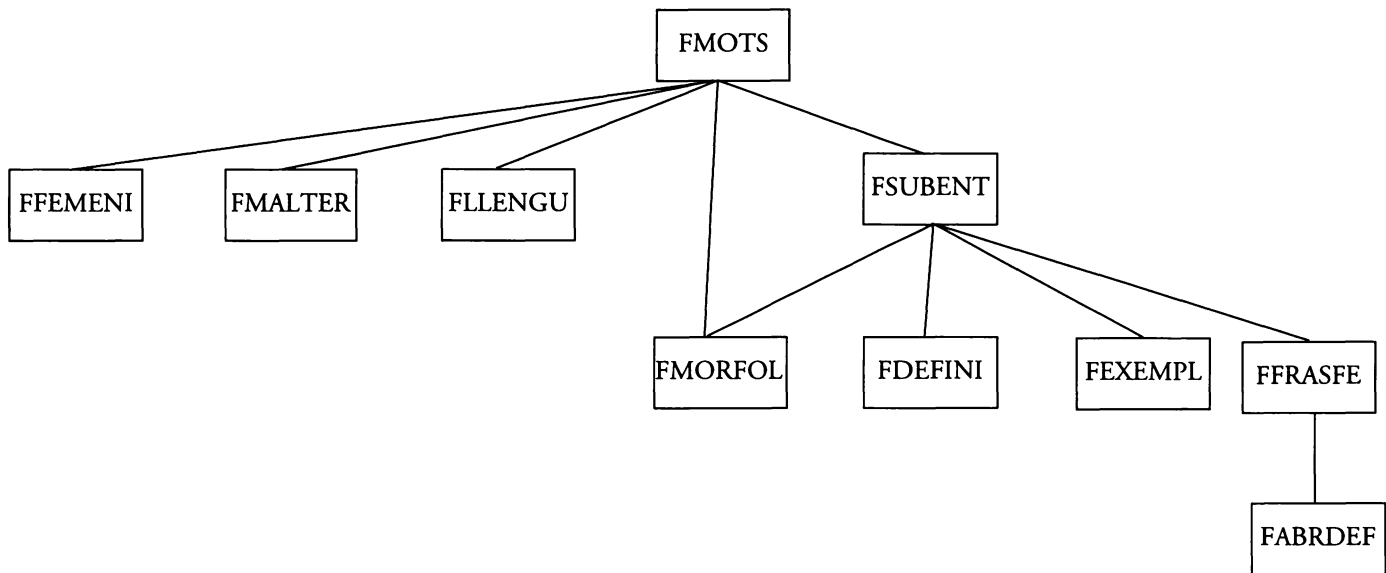


FIGURA 10

Esquema de les relacions entre els arxius i els camps de la BDOL

