

# CORPUS LITERARI DIGITAL: UN PROJECTE DE RECUPERACIÓ, PRESERVACIÓ, DIFUSIÓ I ESTUDI DEL PATRIMONI LITERARI CATALÀ. EINES PER A LA RECERCA

Joan R. VENY-MESQUIDA  
Càtedra Màrius Torres  
Universitat de Lleida

## INTRODUCCIÓ

Aquest treball és el tercer dedicat a l'explicació del projecte del Corpus Literari Digital (CLD) de la Càtedra Màrius Torres de la Universitat de Lleida. Amb el mateix títol de «Corpus Literari Digital: un projecte de recuperació, preservació, difusió i estudi del patrimoni literari català», els dos primers tractaven el que s'hi especificava en els subtítols: «Història i directrius» de la idea, en el primer (Malé / Veny-Mesquida, 2014), i «Descripció i funcionament» del projecte, en el segon (Veny-Mesquida / Malé, 2013). Aquesta tercera entrega vol ser un complement a allò que s'hi exposava en aquells escrits —i per això, per copsar millor el que es desenvoluparà en aquest, en recomano la lectura prèvia—, amb l'objectiu, ara, de separar i explicar amb exemples concrets les eines que ofereix el projecte per a la cerca de materials de cara a la recerca en literatura: concretament, aquesta tercera part pretén explicar la utilització dels diferents camins que ofereix el CLD per buscar títols d'obres i de textos i consultar-ne el contingut i per trobar paraules dins els textos i, també, l'ús dels diferents aplicatius implementats específicament per buscar informació dins les històries de la literatura catalana digitalitzades al Corpus. He maldat per minimitzar la repetició d'algun dels aspectes ja tractats en els dos primers articles, tot referint-m'hi de biaix, sintèticament o, en tot cas, mirant d'aportar nous exemples il·lustradors.

Per començar, caldrà explicar, molt sumàriament, el sentit i l'abast de les paraules clau del títol d'aquest escrit: «recuperació», «preservació», «difusió» i «estudi».

1) *Recuperació*. Un dels avantatges més palesos de la digitalització és que permet reunir, en una sola plataforma, materials literaris de procedència diversa. El procés de recuperació consisteix a cercar aquests materials a biblioteques, arxius i fundacions, tant públics com privats. Per exemple: per les notícies que en tenim, no hi ha cap biblioteca del país que tingui una col·lecció completa (130 números) de la revista *Lleida* (Lleida, 1921 i 1925-1930); per tant, per aconseguir un exemplar de cada número va caldre anar al Fons Romà Sol - Carme Torres de la Universitat de Lleida, al fons de la Biblioteca de

Catalunya, a l'Arxiu Històric de la Paeria i a dues biblioteques particulars, de manera que ara sí que hi ha una biblioteca, ni que sigui virtual, amb la col·lecció completa: la del CLD.

El mateix s'esdevé, és clar, amb les edicions de les obres dels escriptors, la totalitat de les quals difícilment es troba reunida en una sola biblioteca. El CLD, en canvi, permet d'oferir en un sol web totes les obres —i les edicions que calgui de cadascuna—, procedents de biblioteques diferents, de vegades fins i tot de particulars. De Salvador Espriu, per exemple, que ha estat una de les darreres i més integrals incorporacions, s'han digitalitzat edicions procedents de la Biblioteca de Catalunya, del Centre de Documentació i Estudi Salvador Espriu i de biblioteques privades, perquè, en un cas com el seu —que, com és sabut, controlava sempre que podia i fins allà on li era permès qualsevol nova publicació de les seves obres—, calia recollir totes les edicions en què hi ha constància de la seva intervenció, informació que s'ha pogut obtenir a partir d'un buidatge exhaustiu dels volums de l'edició crítica. Cal dir, a tall informatiu, que, per raons diverses, encara no estan disponibles al web totes les edicions digitalitzades.

Encara el mateix es podria dir dels materials autògrafs, però en una proporció molt més elevada, donada la dispersió que aquests solen patir. Per no deixar l'exemple d'Espriu, els seus autògrafs es troben repartits entre *a)* arxius institucionals, com el Centre de Documentació i Estudi Salvador Espriu d'Arenys de Mar, la Casa-Museu Llorenç Villalonga de Binissalem, a Mallorca, l'Arxiu Històric Comarcal de Santa Coloma de Farners, l'Arxiu Històric de l'Ajuntament de l'Alguer o el Archivo General de la Administración de Alcalá de Henares; *b)* fons d'editorials, com Edicions 62, Edicions del Mall o La Polígrafa; *c)* fons personals de biblioteques públiques, com el Fons Jaume Vidal Alcover —Maria Aurèlia Capmany de la Universitat Rovira i Virgili, el del Pare Evangelista Vilanova de la Biblioteca de Montserrat, el de Joan Colomines de l'Arxiu Nacional de Catalunya, els de Tomàs Garcés i Joan Segura de la Biblioteca de Catalunya, el de Ferran de Pol de la Biblioteca Pública d'Arenys de Mar...; i *d)* arxius particulars com, sobretot, el de Sebastià Bonet i el de Ricard Salvat, entre molts d'altres. De tots aquests, el fons més important és, sense cap mena de dubte, el que es conserva a l'arxiu d'Edicions 62 —algun dia caldrà fer l'inventari de tots els materials que aquest arxiu custodia i escriure'n la història—, per la gran quantitat de manuscrits, mecanoscrits, còpies manuscrites o mecanoscrites, fotocòpies amb esmenes autògrafes, exemplars impresos amb anotacions de l'autor, etc. que conté. La concentració d'aquesta suma de materials en aquest fons, gens habitual, s'explica per la dilatada relació de l'autor amb l'editorial, semblantment al que s'esdevé amb l'arxiu particular de Ricard Salvat, que guarda també força materials pels cronològicament allargassats vincles entre tots dos escriptors.

2) *Preservació i difusió.* Inventariats i localitzats els materials, cal procedir a la seva digitalització, que és el nom que rep el pas del document «analògic» al digital i que compta ja amb unes directrius ben marcades (vegeu Biblioteca Nacional de Australia, 2003 i CBUC, 2010). La CMT tracta amb tres grans tipus de documents «analògics»: en paper, sonors i audiovisuals, els quals, al seu temps, poden presentar-se en suports i característiques prou diversos que cal analitzar en cada cas perquè poden exigir tècniques de digitalització diferents. Per exemple: per digitalitzar un manuscrit conservat en fulls solts és indicat un escàner pla, però si el document és un llibre enquadernat caldrà algun tipus d'escàner que no en forci el llom (amb la base en forma de «V» o dividida en dues

parts basculants amb un espai central per al llom); si els fulls transparenten, caldrà disposar una cartolina rere cada full escanejat, etc. El resultat de la digitalització és un arxiu que guarda les característiques del document analògic tal com s'hi presenten. Cal tenir present que l'objectiu de la digitalització és el de preservar el document, l'exemplar analògic concret sotmès al procés, no pas la reconstrucció ideal. Per això els arxius generats en primera instància són els de *preservació*, que es guarden sense cap mena d'intervenció tècnica. Per a aquests primers arxius, els estàndards de digitalització aconsellen el format TIFF (Tagged Image File Format), sense compressió (és a dir sense pèrdua d'informació), amb una resolució de 300 punts per polzada i una profunditat de color de 24 bits (v. *Adobe TIFF Resources Web*: <http://partners.adobe.com/public/developer/tiff/index.html>). Per als documents sonors, la CMT utilitza el format WAV (waveform audio format), sense compressió, estèreo, amb una freqüència de mostratge de 44,1 KHz i 16 bits.

Una còpia d'aquests arxius és sotmesa, quan cal, a un procés de millora que consisteix a retallar les parts sobreres (marges a sang, en el cas d'arxius d'imatge, i principis i finals, en el cas d'arxius de so o de vídeo) i a «netejar», en la mesura del possible, els arxius. En el cas d'arxius d'imatge, la neteja pot consistir en la millora dels contrastos, el retoc del color, etc.; mentre que als arxius de so s'apliquen processos de compressió de so i de reducció de soroll i *hiss*, de *clicks* i *pops*, etc. Els arxius resultants, que serien els que es farien servir per a futures edicions facsímils en paper o en CD, es guarden en format TIFF i WAV, respectivament, sense compressió.

3) D'aquestes còpies es generen les respectives conversions per a la *difusió* en web, de menor qualitat, a fi de reduir-ne al màxim el «pes» i, així, accelerar-ne la consulta: JPG (a 150 dpi, color real) per als documents en paper i MP3 (32 kbps, 44,1 KHz, Mono 22:1:1).

4) *Estudi*. La disposició d'aquests materials a l'abast del gran públic persegueix no tan sols la difusió del coneixement d'aquests materials del patrimoni literari català, sinó també la promoció del seu estudi. En aquest sentit, val la pena de dir aquí que, en l'àmbit acadèmic, hi ha constància que el CLD s'utilitza com a eina de suport a la docència (en l'ensenyament universitari però també en el secundari) i a la recerca, ja que la facilitat de l'accés simplifica i afavoreix l'elaboració de treballs d'investigació. Més enllà, tanmateix, de l'àmbit acadèmic, l'aprofitament d'aquests materials en el camp de la investigació i la recerca és ja un fet. La Càtedra Màrius Torres, com és lògic, n'ha tret especial profit, des de diverses línies d'actuació.

a) Amb l'edició de facsímils, per exemple: el 2006, d'un primer número, de 1932, pràcticament desconegut, de la revista *Art de Lleida*, i de la col·lecció completa (inclòs el número 9, prohibit per la censura quan ja estava en galeres) de la revista *Poemes (Barcelona, 1963-1964)*, el 2010, acompanyat d'un complet estudi de Jaume Aulet.

b) Amb la celebració de reunions científiques dedicades monogràficament a l'estudi dels materials del CLD, per exemple, sobre les revistes dels anys vint i trenta (vegeu Malé / Veny-Mesquida, 2008) i la participació en trobades de recerca amb l'estudi d'exemples concrets d'aquests materials (vegeu Malé / Veny-Mesquida, 2013).

c) Amb l'edició d'estudis monogràfics sobre aquests materials (vegeu Camps / Veny-Mesquida, [en premsa]) i d'edicions crítiques d'aquests materials (la CMT té en projecte l'edició crítica de les obres completes de Màrius Torres, en paper i en línia).

## 1. ESTAT DE LA QÜESTIÓ. EL CORPUS CONSULTABLE

El procés d'integració dels materials que es visualitzen al CLD es realitza en dues fases: *a)* la digitalització i *b)* el registre dins la base de dades SQL que en gestiona la visualització o/i audició i en permet la consulta. Aquesta segona fase constitueix un procés força complex i lent, que demana la màxima atenció durant la seva execució. La base de dades serveix per administrar la informació dels materials digitalitzats: de cada llibre, de cada número de revista, de cada plec de manuscrits, de cada disc o casset o bobina magnetofònica, etc., digitalitzats, es fa l'inventari i buidatge dels textos que conté (i també de les il·lustracions i fotografies), de manera que cada text esdevé un registre bibliogràfic independent, al qual va associat l'arxiu (o arxius) on es troba el text, que pot ser d'imatge (si el text està en un manuscrit, una pàgina de llibre o una revista) o bé de so o de vídeo (si és un enregistrament). D'aquesta manera, en accedir als registres bibliogràfics es pot accedir instantàniament també als documents en qualsevol format (imatge, àudio o vídeo).

Ara bé: no tots els documents que es poden visualitzar o escoltar al web estan ja registrats a la base de dades. Per centrar-nos en una sola secció del CLD, la de l'Hemeroteca, actualment estan completament buidades les publicacions periòdiques següents: *Hèlix*, *La Mà Trencada*, *Monitor*, *La Nova Revista*, *Quaderns de Poesia*, *Quatre coses*, *La Revista dels Llibres*, *Revista de Poesia*, *Rosa dels Vents*, *Teatre català* i *El Ventall del Poeta*. La resta de revistes encara estan en procés d'incorporació a la base de dades, com també part dels llibres de la secció Biblioteca i la pràctica totalitat dels enregistraments de la Fonoteca i tots els de la Videoteca. Això justifica que hi hagi obres o enregistraments —amb els textos que contenen— que es poden visualitzar o escoltar al web però que si es busquen a través del cercador general no apareixeran als resultats. La inversió econòmica —que depèn dels ajuts i de les subvencions— i de temps que aquesta tasca exigeix n'alenteix molt el procés.

## 2. CERQUES A TRAVÉS DEL CERCADOR

Ara bé: el registre a la base de dades dels textos continguts en una revista, una edició, un autògraf o un enregistrament permet d'obtenir-ne l'accés immediat, de formes diverses. La més completa, perquè fa possible una combinatòria de paràmetres més complexa, és sense dubte a través del cercador, que permet fer quatre tipus de consultes:

«Cerca de textos» i «Cerca d'obres». Els dos primers tipus de cerques atenyen els materials sobre els quals es durà a terme la cerca: sobre els títols de les obres o sobre els dels textos que aquelles contenen. En ambdós casos, per materialitzar la cerca sols cal «interpretar» el que diu la pàgina del cercador d'aquesta manera:

A la secció	Totes Hemeroteca Biblioteca Manuscrits Fonoteca Videoteca	el camp	Qualsevol Autor del text Curador del text Traductor del text ...	Conté el/s mot/s Conté la seqüència Coincideix amb Comença per	[text a cercar]
-------------	--	---------	--	---	-----------------

És a dir que, per exemple, per buscar els textos publicats en revistes per tots els autors els cognom dels quals contingui la seqüència «Carner», caldrà triar les opcions marcades en negreta:

A la secció	Totes <b>Hemeroteca</b> Biblioteca Manuscrits Fonoteca Videoteca	el camp	qualsevol <b>autor del text</b> curador del text traductor del text ...	conté el/s mot/s <b>conté la seqüència</b> coincideix amb comença per	<b>Carner</b>
-------------	---	---------	---	--	---------------

El text a buscar es pot sotmetre a quatre tipus de coincidències: «Conté el/s mot/s» significa que es llistaran els registres el camp especificat dels quals contingui els mots entrats en la cerca *en qualsevol lloc i ordre*; «Conté la seqüència» implica que s'enumeraran els registres el camp especificat dels quals contingui, en alguna part, *la seqüència exacta* de mots entrats en la cerca; «Coincideix amb» vol dir que es relacionaran els registres el camp especificat dels quals coincideixi *completament* amb la seqüència de mots entrada en la cerca; i «Comença per» indica que es llistaran els registres el començament del camp especificat dels quals coincideixi exactament amb la seqüència de mots entrada en la cerca.

Els resultats d'aquestes consultes es poden afinar encara més si es combinen amb les tres «Opcions de cerca» que ofereix el sistema. Als registres de la base de dades del CLD es transcriuen els noms dels autors, dels títols, de les editorials, etc., tal com estan escrits en el document original però es fan correspondre amb un de sol (que sol ser el «definitiu»), de manera que les diferents formes amb què un autor pot signar els seus textos (J. V. Foix, J-V Foix, J. V. F., Fòcius, Ramon N. Giralt, Josep del Mas, etc.) estan relacionats a través del seu nom «real»: Josep Vicenç Foix i Mas. L'aplicatiu del CLD, per tant, ha de donar l'opció a l'usuari de poder triar sobre quina d'aquestes formes vol que es realitzi la cerca. D'aquí les tres «Opcions de cerca» a les quals em referia al principi d'aquest paràgraf:  $\alpha$ ) «Llistar les formes concurrents i preguntar»;  $\beta$ ) «Trobar totes les formes concurrents d'aquesta cerca»;  $\gamma$ ) «Trobar sols les ocurrencies exactes d'aquesta cerca».

Suposem que l'usuari busca textos de l'autor «Riba, Carles» o obres publicades per l'editorial «Renaxensa». Si tria l'opció  $\alpha$ , el sistema buscarà totes les formes associades a «Riba, Carles» (Riba-Bracons, Carles; Riba, C.; Jordi March; etc.) o a «Renaxensa» (Renaixença, Renaxensa, Renaixensa, etc.) i les ensenyarà en una llista perquè l'usuari pugui triar la que li interessa (o totes); en canvi, si selecciona l'opció  $\beta$ , el sistema presentarà els resultats relatius a totes les formes de «Riba, Carles» o de «Renaxensa», directament, sense l'opció d'haver de fer cap tria; mentre que si escull l'opció  $\gamma$ , el sistema presentarà sols els resultats relatius a les ocurrences exactes de «Riba, Carles» i «Renaxensa», obviant, doncs, els resultats de les altres formes (Riba-Bracons, Carles; Riba, C.; Jordi March; o Renaixença, Renaixensa, etc.).

«Cerca guiada» i «Cerca avançada». Els altres dos tipus de cerques atenyen sobretot el procediment d'obtenció de resultats, més que no pas els resultats mateixos. La cerca guiada proporciona dreceres per a les cerques més habituals, està construïda com un diàleg entre l'usuari i el motor de cerca del CLD, de manera que la interfície va acompanyant l'usuari en cadascun dels passos que dona, i va canviant en funció de l'opció que vagi seleccionant consecutivament. Per exemple, per fer la mateixa cerca esmentada en l'apartat anterior («textos publicats en revistes per tots els autors el cognom dels quals contingui la seqüència «Carner»»), caldrà triar (sense especificar dates):

<b>Cercar</b>	a totes les seccions <b>a l'Hemeroteca</b> a la Biblioteca a la secció Manuscrits a la Fonoteca a la Videoteca	obres <b>textos</b> versions	<b>de</b> a cura de traduïts per publicats a publicats per de la collecció	<b>Carner</b>	publicats entre [dd/mm/aaaa]- [dd/mm/aaaa]
---------------	--	------------------------------------	---	---------------	--

I aleshores el sistema preguntarà si l'usuari vol llistar totes les formes concurrents de «Carner» o prefereix triar-ne una de les trobades («Carner, Josep» o «Carner-Ribalta, J.»). En canvi, per cercar les diferents versions de la primera Elegia de Bierville de Riba la tria serà aquesta:

<b>Cercar</b>	<b>a totes les seccions</b> a l'Hemeroteca a la Biblioteca ...	obres <b>textos</b> <b>versions</b>	<b>del text</b>	<b>Era secret</b>	publicats entre [dd/mm/aaaa]- [dd/mm/aaaa]
---------------	---	---	-----------------	-------------------	--

La cerca avançada, finalment, ve a ser com una cerca simple d'obra o de text però amb la possibilitat de combinar diversos camps i d'acotar els resultats per dates. En l'exemple següent, es llistaran els pròlegs de Carles Riba a obres de Poe publicades entre 1910 i 1955:


A la secció <b>Biblioteca</b>	
Gènere del text	Conté el/s mot/s: Introducció
Autor del text	Conté el/s mot/s: Riba
Autor de l'obra	Conté el/s mot/s: Poe
<input type="button" value="+"/> <input type="button" value="-"/>	Data entre 31/01/1910 i 31/12/1955
<input type="button" value="Cercar"/>	

En aquest altre, el resultat serà la llista de ressenyes que Marià Manent va publicar a la *Revista de Poesia*:

A la secció <b>Hemeroteca</b>	
Autor del text	Conté el/s mot/s: Manent
Títol de l'obra o revista	Conté el/s mot/s: Revista de Poesia
Gènere del text	Conté el/s mot/s: Ressenya
<input type="button" value="+"/> <input type="button" value="-"/>	Data entre dd/mm/aaaa i dd/mm/aaaa
<input type="button" value="Cercar"/>	

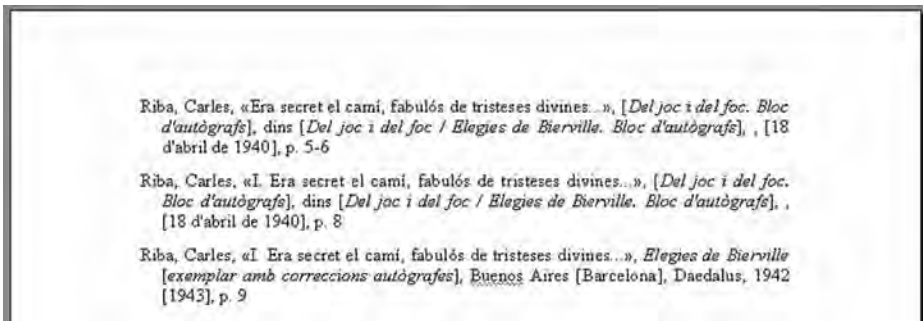
Val a dir, per tancar aquest punt, que està prevista la implementació, al Cercador, de l'opció de buscar una paraula o seqüència de paraules dins el text dels documents digitalitzats. Aquesta funcionalitat serà possible perquè els materials digitalitzats estan essent sotmesos a un procés de reconeixement òptic de caràcters (OCR) que converteix les imatges dels impresos en text.

### 3. CERQUES EN REVISTES I LLIBRES

El cercador, tanmateix, no és l'única forma d'accedir als materials registrats a la base de dades del CLD: actualment, bona part dels documents impresos (revistes i llibres, de l'Hemeroteca i la Biblioteca respectivament: més endavant la possibilitat s'ampliarà a la resta de seccions) disposen d'un sumari que el sistema és capaç d'elaborar a partir de la base de dades. Aquest sumari es pot consultar de forma individualitzada, per a cada número d'una determinada revista o per a cada obra editada, o globalment, per a tots els números de la revista o per a totes les obres llistades a la pàgina principal de cada autor. Al sumari individual s'accedeix a través de la icona  situada al costat de cada exemplar (revista o obra) i al global a través del menú «Cerques i índexs» disposat al principi de la pàgina principal de cada revista o de cada autor.

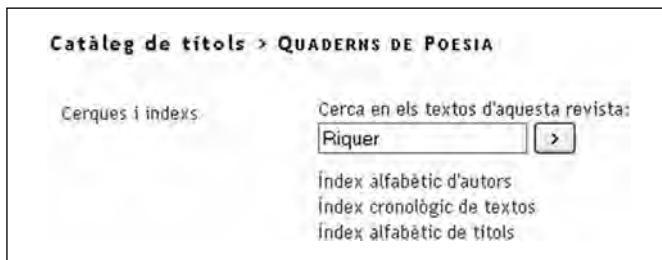
D'aquesta manera, hom pot disposar del sumari dels títols dels textos de qualsevol número de revista o de qualsevol obra i, més encara, al sumari de tots els números d'una

revista o de tots els textos d'un autor, i pot accedir a cada text amb un sol clic de ratolí. Per a una major cobertura de les necessitats de l'usuari, el sumari es pot presentar per ordre alfabètic de títol de text o cronològic de data i, a través de l'opció «Exportar a RTF», el sumari es converteix en un document RTF (Rich Text Format) que s'obre automàticament amb el processador de textos que l'usuari tingui instal·lat en el seu ordinador, perquè es pugui manipular a voluntat o copiar i enganxar dins un altre document. El document RTF conté la llista de resultats de la cerca en forma de referències bibliogràfiques «estàndards»:



Més endavant s'implementarà la possibilitat de triar altres formats o, fins i tot, que l'usuari pugui fixar-ne el/s propi/s.

En un altre ordre de coses, val a dir, quant a la cerca en revistes i llibres, que el CLD ja permet buscar paraules dins el cos d'algunes de les revistes i dels llibres. Concretament, en les seccions d'Hemeroteca i Biblioteca, això és factible sempre que en la pàgina principal d'una revista o d'un autor apareix l'opció «Cerques i índexs». Si el lector busca «Riquer» en l'apartat corresponent de la pàgina principal dels *Quaderns de Poesia*



o «pluja» en la de B. Rosselló-Pòrcel, comprovarà que el resultat no és tan sols una rastellera d'imatges de les pàgines on apareix la paraula en qüestió sinó la llista dels textos que la contenen, amb la seva referència completa i, entre claudàtors, la pàgina concreta on surt el mot cercat:





#### 4. APLICATIUS ESPECÍFICS

Amb l'objectiu d'abastar el major nombre possible de les necessitats dels investigadors en literatura contemporània, el web del CLD incorpora, a banda de les pàgines descrites fins aquí, aplicatius específics per accedir a materials concrets, bona part dels quals no formen pròpiament part del Corpus. Aquest és el cas, per exemple, dels aplicatius que gestionen el *Fitxer Espriu* ([http://www.catedramariustorres.udl.cat/materials/manuscrits/\\_autors/espr/fitxer/](http://www.catedramariustorres.udl.cat/materials/manuscrits/_autors/espr/fitxer/)), el *Diccionari de publicacions periòdiques* (<http://www.catedramariustorres.udl.cat/materials/hemeroteca/>) o les *Preceptives poètiques*, que ja han estat descrits en un altre treball (vegeu Veny-Mesquida / Malé, 2013). Aquí em volia centrar en la pàgina destinada a la gestió de les *Històries de la literatura*, per la seva utilitat, i perquè, si a Veny-Mesquida / Malé (2013) era presentat com un projecte, ara és ja una realitat, bé que incompleta.

Com la resta de pàgines dedicades a aquests materials singularitzats, la de les *Històries de la literatura* està estructurada en dues parts. A la superior, es presenten les pàgines de les obres digitalitzades en una llista que es desplega en clicar damunt «Consulta dels textos». Val a dir que en aquest inventari, apareixen obres que no s'ocupen de tots els segles (la *Historia del Renacimiento literario contemporáneo...* de Tubino, de 1880) o de tots els gèneres (la *Renaissance de la Poésie Catalane* d'Albert Savine, 1887) o de tot el territori (*L'aportació de l'Occident català...* de Ramon Xuriguera, de 1936), però que han estat, pel seu ressò o la seva rellevància en el moment de la seva publicació, i són, pel seu estatus de «clàssics» en el seu àmbit, útils. La llista és exhaustiva, i un sol clic damunt les que ja estan disponibles permet visualitzar-les i llegir-les com si l'usuari tingués el llibre a les mans.

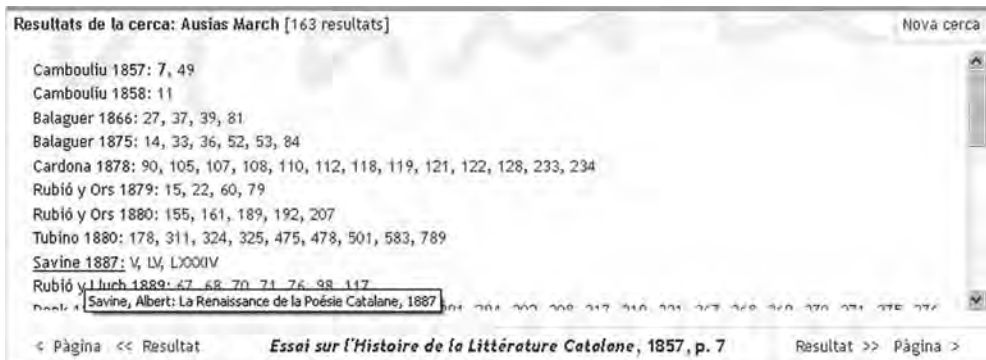
En la part inferior de la pàgina hi ha un aplicatiu que permet cercar qualsevol paraula o seqüència de paraules dins el text d'una història de la literatura en concret o de totes si no se'n tria cap. Si, per posar un exemple més que il·lustre, hom busca «Ausias March» (recordem que els cercadors del CLD no discriminen els accents):

Consultes al cercador

Cercar:

A:

el resultat serà una llista com aquesta:



amb les obres ordenades cronològicament i abreujades pel sistema d'autor i data (en posar-hi el ratolí al damunt apareix el títol sencer, com es pot veure en la imatge de l'exemple amb *La Renaissance de la Poésie Catalane* d'Albert Savine) i seguides de les pàgines on apareix «Ausias March»: en clicar damunt cada pàgina el visor en mostra la imatge corresponent. Cal tenir present que la cerca es fa sobre textos elaborats mitjançant un reconeixement òptic de caràcters (OCR) automàtic, que té un nivell de precisió directament proporcional a la qualitat de l'original: com més net de taques (de tinta, de rovell, d'humitat) estigui el paper i més bona sigui la impressió de la lletra, més ajustat a l'original resultarà el text reconegut òpticament. Això vol dir que fins que no es puguin repassar i col·locar manualment aquests textos creats de forma automàtica amb els originals impresos per esmenar-los, cal tenir clar que els resultats no seran mai del tot fiables.

## FINAL

La finalitat d'aquest escrit no era altra que destacar les eines que ofereix el CLD que habitualment poden fer servei a la investigació en literatura. Tanmateix, la millor manera d'entendre i de copsar el seu sentit i el seu abast és experimentar-hi, explorar-ne les múltiples possibilitats que presenta el web. Animo el lector a passejar-s'hi i comprovar-ne la utilitat.

## BIBLIOGRAFIA

- Biblioteca Nacional de Australia 2003: Biblioteca Nacional de Australia, *Directrices para la preservación del patrimonio digital*, en línia: <http://unesdoc.unesco.org/images/0013/001300/130071s.pdf> [consulta: maig 2014]
- CAMPS I ARBÓS, Josep / VENY-MESQUIDA, Joan R. (en premsa): *L'Arxiu Sonor de Poesia de Joan Colomines i Puig*. Lleida: Càtedra Màrius Torres; Pagès Editors.
- CBUC 2010: *CBUC: Estàndards de digitalització: requeriments mínims* (actualitzat a octubre 2010), Barcelona, CBUC, 2010; en línia: [http://www.recercat.cat/bitstream/handle/2072/97495/1010EstandardsRequerimentsMinims\\_Public.pdf](http://www.recercat.cat/bitstream/handle/2072/97495/1010EstandardsRequerimentsMinims_Public.pdf) [consulta: maig 2014]
- MALÉ, Jordi / VENY-MESQUIDA, Joan R. (ed.) (2008): *Les revistes literàries a la Catalunya d'entreguerres. Crítica, recepció, traducció*, Lleida: Aula Màrius Torres; Pagès Editors.
- MALÉ, Jordi / VENY-MESQUIDA, Joan R.: «La recerca en el Corpus Literari Digital: dos exemples», p. 183-196, dins BACARDÍ, M. / FOGUET, F. / GALLÉN, E. (ed.) (2013): *La literatura catalana contemporània: intertextos, influències i relacions*. Barcelona: Institut d'Estudis Catalans.
- MALÉ, Jordi / VENY-MESQUIDA, Joan R. (2014): «Corpus Literari Digital: un projecte de recuperació, preservació, difusió i estudi del patrimoni literari català. Història i directrius», dins FÀBREGAS, Immaculada (ed.): «V Colloque International de l'Association Française des Catalanistes: langue, littérature et civilisation 13-15 octobre 2011», *REC - Revista d'Estudis Catalans* 1, en premsa.
- VENY-MESQUIDA, Joan R. / MALÉ, Jordi (2013): «Corpus Literari Digital: un projecte de recuperació, preservació, difusió i estudi del patrimoni literari català. Descripció i funcionament», dins CARRERAS, M. / PUIGDEVALL, N. / RIGOBON, P. / RIPA, V. (ed.): *La ciutat de l'amor. Scrivere la città, raccontare i sentimenti. 23-25 febbraio 2012*, Alessandria: Edizioni dell'Orso, p. 427-442.