

Sobre la riquesa lèxica de Verdaguer: una primera prospecció quantitativa

Jordi Ginebra
(Universitat Rovira i Virgili)

Introducció

El treball que presento en aquesta comunicació és el fruit alhora d'un neguit i d'una sensació d'impotència. El neguit prové de la convicció que no n'hi ha prou que entre els estudiosos es repeteixi l'afirmació que l'obra de Verdaguer és d'una gran riquesa lèxica, sinó que cal, a més, mostrar objectivament aquesta riquesa i determinar-ne l'abast. La sensació d'impotència neix de percebre la magnitud de l'empresa: determinar objectivament en què consisteix i fins on arriba el valor del lèxic verdaguerià només serà possible quan, prèviament, se n'hagi confeccionat l'inventari lexicogràfic. I no es podrà confeccionar l'inventari lexicogràfic de l'obra de Verdaguer —el diccionari— fins que no s'hagi conclòs el projecte d'edició de les seves obres completes.

Mentrestant el neguit té més força que la sensació d'impotència, i això és el que m'ha dut a fer un primer assaig d'exploració quantitativa per mirar de determinar objectivament la riquesa lèxica de l'obra de Verdaguer. Naturalment, és una anàlisi feta amb un mètode específic, i no és doncs l'única mena d'anàlisi possible ni probablement la més conclouent. En qualsevol cas, serà complementària d'altres que també caldria fer.

Uns quants aclariments i comentaris sobre una primera recerca sense resultats

Abans de posar-se a treballar en l'estudi que ara presento calia mirar d'objectivar de manera precisa i senzilla —sense necessitat d'entrar en disquisicions teòriques— què és el que volem dir quan parlem de *riquesa lèxica*. El concepte, si no el matisem, és essencialment quan-

titatiu. Si establim l'inventari de mots usats per un autor, llavors l'obra d'aquest autor tindrà més riquesa lèxica com més extens sigui aquest inventari. Com que el diccionari dels mots usats per Verdaguer encara no existeix, aquesta quantificació no es pot fer. Ara: si tinguéssim l'inventari dels mots de Verdaguer podríem dir quin nombre d'unitats lèxiques conté la seva obra, però la xifra seria encara poc significativa si no la podíem comparar amb la d'altres autors. Per aquest motiu la feina per fer és de grans dimensions. Un cop tinguem l'inventari dels mots usats per uns quants autors, podrem dir que el lèxic de l'autor A és més ric que el lèxic de l'autor B si el seu inventari lexicogràfic és més extens.

Tot i així, aquesta quantificació comparada de la riquesa lèxica de dos o més autors podria igualment ser poc significativa si l'extensió de l'obra de l'un i de l'altre no és comparable. Hi ha una llei general que estableix que, com més extens és un text, més nombre d'unitats lèxiques conté. Així, si un text de 200.000 paraules (ocurrències) pot tenir 10.000 unitats lèxiques diferents, un text de 20.000 paraules és molt difícil que arribi a les 5.000 unitats lèxiques diferents. Si apliquem aquesta llei a la comparació de dos autors amb una obra completa molt desigual quant a l'extensió, el que en resultarà és que l'autor amb l'obra més extensa serà sempre el que aparegui amb un lèxic més ric.

Per mirar de corregir aquest resultat es pot tenir en compte la riquesa lèxica *relativa*, és a dir, la proporció entre extensió de l'obra i nombre d'unitats lèxiques diferents. D'aquesta proporció en pot sortir una *ratio* que resultaria de dividir el nombre d'unitats lèxiques diferents d'una obra o d'un text pel nombre d'ocurrències totals de l'obra o del text. Així, es podria considerar aquesta *ratio* indicadora de la riquesa lèxica relativa d'un autor o d'un text, i determinar que la riquesa lèxica relativa és més significativa que la riquesa lèxica absoluta a l'hora de valorar la riquesa lèxica de l'obra d'un autor.

Però si es comparen els índexs de riquesa lèxica relativa de diferents textos segons la *ratio* que acabem d'establir, el resultat és que aquest índex és, en general, més baix com més extens és el text. Dit d'una altra manera: un text més llarg conté un nombre més alt d'unitats lèxiques diferents que no pas un text més curt, però *proporcionalment a l'extensió* un text curt conté més unitats lèxiques diferents que no pas un text llarg. Això vol dir que la riquesa lèxica relativa, si no volem introduir en la *ratio* factors de

correcció relacionats amb l'extensió, només és significativa si comparem textos o obres de llargada equiparable.

Després d'haver fet les reflexions que acabo de detallar vaig proposar-me estudiar quina era la riquesa lèxica relativa d'alguna de les obres de Verdaguer. Aquest estudi era possible gràcies a l'existència del Corpus Textual Informatitzat de la Llengua Catalana, de l'Institut d'Estudis Catalans (des d'ara Corpus o CTILC), que, com se sap, està constituït pel text, en suport digital, d'obres literàries i no literàries publicades entre els anys 1833 i 1988. El Corpus conté aproximadament 3.300 textos diferents, que sumen més de 50 milions d'ocurrències. El Corpus —que es pot consultar en línia a través del Portal de la llengua de l'Institut d'Estudis Catalans— també ofereix una fitxa per a cada una de les obres, una fitxa que informa —pel que ara ens interessa— del nombre d'ocurrències i del nombre de lemes (unitats lèxiques) de cada obra.

El meu propòsit era molt senzill, per bé que lent de realització. Es tractava d'agrupar les obres en blocs segons l'extensió, obtenir l'índex de riquesa lèxica relativa de cada obra dividint, com he dit, el nombre de lemes pel nombre d'ocurrències, i establir el resultat en forma d'una llista jerarquitzada per a cada bloc. D'aquesta manera podria obtenir el rànquing de riquesa lèxica relativa de les obres de cada bloc i comprovar, així, quina era la posició de les obres de Verdaguer que figuren al Corpus.

La lentitud del procés em va fer desistir, de moment, d'examinar les fitxes de tot el Corpus, i em vaig limitar a analitzar les fitxes de les obres d'una de les dues seccions en què està dividit el Corpus, el subcorpus literari (l'altra secció és el subcorpus no literari, que conté obres que no són de literatura). El subcorpus literari conté 23.105.591 ocurrències (una mica menys de la meitat del Corpus) i 1.011 títols. De Verdaguer, hi ha recollides 12 obres: *Dos màrtirs de ma pàtria, o siga Lluçà i Marçà* (1865), *Romanç llegit a l'ombra de la font del Desmai* (1869), *Passió de Nostre Senyor Jesucrist* (1873), *L'Atlàntida* (1877), *A Barcelona* (1883), *Canigó* (1886), *Excursions i viatges* (1887), *Dietari d'un pelegrí a Terra Santa* (1889), *Jesús Infant. Nazareth* (1890), *Roser de tot l'any* (1894), *Flors del Calvari* (1896) i *Al Cel* (1903).

El resultat d'aquest estudi no va ser particularment engrescador, almenys en relació amb l'objectiu que tenia, que era atènyer uns primers indicis de demostració objectiva de la riquesa lèxica de Verdaguer. Si bé totes les obres de Verdaguer tenen un índex de riquesa lèxica relativa que

és per damunt de la mitjana del bloc corresponent, no ocupen pas les primeres posicions. Així, l'obra *Excursions i viatges*, que pertany al bloc de textos que tenen entre 30.001 i 40.000 ocurrences, que són 62, ocupa el lloc número 26. En el bloc d'obres que presenten entre 20.001 i 30.000 ocurrences, que són 114, hi ha *L'Atlàntida*, que ocupa el lloc número 24, i *Canigó*, que ocupa el lloc número 53. Només *Dos màrtirs de ma pàtria*, del bloc d'obres d'entre 6.001 i 7.000 ocurrences, pot dir-se que ocupa una posició especialment rellevant: de 55 obres ocupa el lloc número 6.

Repensant l'anàlisi

Els resultats que acabo de presentar em van deixar profundament abatut. Havia estat unes quantes setmanes fent recomptes i divisions i establint mitjanes amb la il·lusió de presentar-me al VI Col·loqui Verdaguer i proclamar públicament que tenia la demostració matemàtica de la gran riquesa lèxica de l'obra de Verdaguer, i havia fracassat. Llavors, després de fer les reflexions que ara resumiré, vaig pensar que havia d'assajar una altra via.

Les meves reflexions en aquell moment em van dur a fer la hipòtesi que el que jo havia examinat fins llavors, que és el que podríem denominar la riquesa lèxica bruta d'una obra, potser no era un element tan significatiu en relació amb l'aportació efectiva de l'obra al cabal lèxic del català contemporani com altres variables també relacionades amb el lèxic. I que potser la valoració de la riquesa lèxica d'un autor no pot fer-se exclusivament basant-se en paràmetres quantitius, sinó que també ha de prendre en consideració factors qualitatius. M'explicaré amb un exemple simple. Pot ser que una obra literària A sigui més rica que una obra literària B en termes estrictes de riquesa lèxica relativa. Però si l'obra A conté una proporció molt més alta de castellanismes que l'obra B, i l'obra B, per tant, conté un nombre més alt d'unitats lèxiques genuïnes, llavors potser cal tenir en compte aquesta circumstància a l'hora de procedir a l'avaluació del lèxic de totes dues obres (almenys si el que es pretén en definitiva, que és el que pretenem quan parlem del lèxic de Verdaguer, és referir-nos a la fixació del lèxic literari del català contemporani). Podríem pensar en altres supòsits interessants: les obres A, B, C i D són semblants quant al seu índex de riquesa lèxica relativa, però les obres A, B i C com-

parteixen en termes generals un mateix tipus de lèxic, mentre que l'obra D destaca perquè el seu lèxic és qualitativament diferent del de les altres tres considerades en conjunt.

Una nova recerca

Com a resultat d'aquestes reflexions vaig idear una altra mena d'anàlisi, que pretenia, doncs, mesurar una nova magnitud. Una magnitud que, si no era ja la riquesa lèxica relativa de Verdaguer —el que he denominat riquesa lèxica bruta—, podria considerar-se que és un tipus de riquesa lèxica *qualificada*. Es tractava d'avaluar, utilitzant igualment les dades del Corpus de l'IEC, el nombre d'unitats lèxiques que Verdaguer havia aportat a la llengua literària contemporània. El nombre i, en una segona fase, la qualitat. No pretenia pas fer-ne un estudi complet sinó un estudi representatiu: una anàlisi parcial que proporcionés dades que poguéssim considerar indicis de l'aportació global de Verdaguer. Com abans, també ara m'interessava la referència contrastiva: si Verdaguer havia contribuït a la formació del lèxic modern amb un nombre més gran d'unitats que altres autors amb els quals se'l pogués comparar.

Aquesta avaluació pot mirar de fer-se per diverses vies, i en vaig triar una. Vaig partir del supòsit (un supòsit merament de treball) que el lèxic del català escrit contemporani és el lèxic que conté el subcorpus literari de l'esmentat Corpus de l'IEC. Aquest subcorpus recull 23.105.591 ocurrences i 77.148 unitats lèxiques (*lemes principals*, d'acord amb la terminologia utilitzada al Corpus). Vaig establir igualment, també merament com a postulat metodològic, que un mot *a* és una aportació de l'autor A al lèxic escrit del català contemporani si, segons el subcorpus literari del CTILC, apareix documentat per primera vegada en una obra de l'autor A. Llavors vaig procedir a la selecció d'una mostra representativa d'unitats lèxiques del subcorpus. El procediment havia de consistir a realitzar, per a cada una de les unitats seleccionades, la consulta pertinent. La consulta es va fer en línia, a través del Portal de la llengua, per mitjà dels filtres que permeten obtenir la informació relativa a la data i l'obra en què cada lema és documentat per primera vegada.

Lògicament, la mostra de lèxic havia de ser completament aleatòria. Vaig decidir fixar la mostra en 3.500 unitats. Per raons tècniques de pro-

cediment, vaig rebutjar els lemes amb més de 100 ocurrences. (El servidor del Portal de la llengua no proporciona mai més de 100 ocurrences en una consulta, i a mi m'interessava, per estalviar temps, no haver de fer més d'una consulta per lema. A més, volia accedir a totes les ocurrences de cada lema examinat, perquè era l'única garantia que l'aplicació informàtica em permetria conèixer la que cronològicament és la primera, ja que, si la consulta afecta lemes de més de 100 ocurrences, l'aplicació te'n «retorna» només una mostra aleatòria.) La limitació quant al nombre d'ocurrences dels lemes de la mostra era també convenient per una altra raó: en un corpus d'obres literàries ordenades cronològicament, les paraules més freqüents de l'idioma apareixen en primer lloc a les obres més antigues, que, lògicament, són les que ocupen els primers llocs de la llista cronològica. Per aquest motiu, les obres més antigues, si se'm permet l'expressió, «tenen avantatge» a l'hora d'obtenir la distinció de ser les que primer documenten una unitat lèxica. Reduir les ocurrences dels mots de la mostra era, així, introduir un factor de correcció parcial en relació amb aquest avantatge.

La mostra de 3.500 paraules, així, va quedar constituïda pels elements següents. Seguint l'ordre alfabètic, els primers 700 lemes de 100 ocurrences del subcorpus, els 700 primers lemes de 50 ocurrences, els primers 700 lemes de 25 ocurrences, els primers 700 lemes de 10 ocurrences i els primers 700 lemes de 5 ocurrences. En el cas que per a algun dels grups no hi hagués al subcorpus 700 lemes amb el nombre d'ocurrences requerit, es va continuar amb els lemes amb nombre d'ocurrences immediatament inferior. (Això va passar amb el grup de lemes de 100 ocurrences, amb el de 50 i amb el de 25.)¹

Encara que el propòsit inicial era avaluar amb aquest mètode les aportacions a la llengua literària contemporània de totes les obres de Verdaguer recollides al subcorpus, per problemes de temps vaig limitar la recerca a tres obres. Les obres que em vaig proposar avaluar van ser *Atlàntida*, *Canigó* i *Dos màrtirs de ma pàtria*. Les dues primeres m'inte-

1. La llista completa de les unitats que formen la mostra va confeccionar-se seguint les dades que figuren al *Diccionari de freqüències*. 2 *Llengua literària* (Joaquim Rafel, director, 1998). Formen part d'aquesta llista els lemes que apareixen a les p. 436-443 (des d'«amansir» fins a «negligència»), p. 477-484 (des d'«abassegador» fins a «melodramàtic»), p. 527-533 (des d'«abarcar» fins a «trabocar»), p. 603-609 (des d'«abacial» fins a «pedrerar») i p. 669-675 (des d'«a» fins a «culblanc»).

ressaven en qualitat d'obres supremes de l'opus verdaguerià, i la tercera pel fet que és el primer poema llarg de joventut de Verdaguer.²

Per tal que l'anàlisi tingués caràcter comparatiu vaig seleccionar altres obres literàries recollides al subcorpus literari del CTILC. Aquestes obres havien de ser de la mateixa època que les obres de Verdaguer. Si eren massa anteriors comptarien amb un avantatge temporal excessiu. I si eren massa posteriors estarien en situació clara de desavantatge. A més, per les raons adduïdes abans, havien de ser d'extensió equiparable a les obres de Verdaguer. Així, el poema *Dos màrtirs de ma pàtria* conté 6.287 ocurrences i, per tant, l'havia de comparar amb obres que tinguessin entre 6.000 i 7.000 ocurrences, d'acord amb els blocs que havia establert per a la recerca de què he parlat abans. El poema *L'Atlàntida* registra 23.229 ocurrences i *Canigó*, 28.351. Totes dues obres pertanyen, doncs, al bloc d'obres que sumen entre 20.000 i 30.000 ocurrences. Vaig buscar, doncs, un conjunt d'obres per comparar-les amb *Dos màrtirs de ma pàtria* i un altre per comparar-les amb *L'Atlàntida* i *Canigó*. Per a la primera comparació vaig seleccionar *Déu i lo món* (1854), de Benet Altet (6.138 ocurrences); *Celísties* (1866), de Francesc Ubach i Vinyeta (6.460 ocurrences), i *Poesies fantàstiques en mallorquí* (1852), de Tomàs Aguiló i Forteza (6.007 ocurrences). Per a la segona comparació vaig seleccionar *L'escanyapobres* (1884) i *Croquis del natural* (1879), de Narcís Oller (22.548 i 21.295 ocurrences, respectivament); *Eures* (1882), d'Antoni Careta i Vidal (20.177 ocurrences); *La pubilla del mas de Dalí* (1866) i *La família del mas dels Salzes* (1880), de Gaietà Vidal i de Valenciano (22.323 i 23.036 ocurrences, respectivament); *Cartes a la dona* (1877), de Rossend Arús (27.711 ocurrences); *Plançons* (1882), de Francesc Fayos (23.874 ocurrences); *Lo pagès de l'Empordà* (1875), d'Antoni Ferrer i Codina (21.605 ocurrences); *Narracions i llegendes* (1874), de Maria del Pilar Maspons i Labrós (28.106 ocurrences); *Lo prodigi del segle* (1868), de Josep Barberà i Canturri (28.959 ocurrences), i *Versos catalans* (1893), de Josep Martí i Folguera (21.531 ocurrences).

Com he dit, el procés de treball va consistir a fer una consulta en línia a través del Portal de la llengua per a cada un dels 3.500 lemes de la mostra, examinar per mitjà dels filtres pertinents si la primera documentació del lema pertany a una de les obres seleccionades i fer la llista i el recomp-

2. Per a les edicions més actuals i rigoroses d'aquestes obres, i també per a valoracions i anàlisis relatives a la llengua que presenten, vegeu VERDAGUER (1995a, 1995b i 2002).

te final de lemes que són documentats per primera vegada en cada una de les obres seleccionades.

Resultats de la recerca

Aquesta vegada el resultat va ser més encoratjador que l'anterior. Recordem que els postulats metodològics de l'estudi, com he dit, eren els següents: a) el subcorpus literari del CTILC recull el cabal lèxic escrit del català contemporani; b) un autor aporta un mot al cabal lèxic escrit del català contemporani si el subcorpus documenta el mot per primera vegada en una obra d'aquest autor.

Com era previsible, la major part de les unitats lèxiques examinades no són protodocumentades en cap de les obres seleccionades. N'hi ha que són documentades per primera vegada en obres anteriors i n'hi ha de protodocumentades en obres posteriors. Del conjunt de les 3.500 unitats, n'hi ha 211 de documentades per primer cop en alguna de les obres de la llista. D'aquestes 211, n'hi ha 28 que són protodocumentades en obres del primer grup (el grup d'obres amb 6.000-7.000 ocurrències) i 183, en obres del segon grup (el grup d'obres amb 20.000-30.000 ocurrències).

Del grup de les quatre obres amb 6.000-7.000 ocurrències, l'obra de Verdaguer, *Dos màrtirs de ma pàtria*, presenta 15 mots documentats per primer cop al subcorpus. Les altres obres en presenten 7 (*Celísties*), 4 (*Poesies fantàstiques en mallorquí*) i 2 (*Déu i lo món*). D'acord amb els supòsits previs establerts, doncs, d'aquestes obres, la de Verdaguer és la que fa més aportacions al lèxic escrit del català contemporani.

Del grup de les tretze obres amb 20.000-30.000 ocurrències, *L'Atlàntida* és la que més unitats documenta per primera vegada: 40. Les altres obres en presenten 36 (*Lo prodigi del segle*), 21 (*La pubilla del mas de Dalt*), 17 (*Canigó*), 16 (*L'escanyapobres*), 14 (*Cartes a la dona*), 9 (*Croquis del natural i Lo pagès de l'Empordà*), 8 (*Narracions i llegendes*), 6 (*Plançons*), 3 (*Versos catalans*) i 2 (*Eures i La família del mas dels Salzes*). D'acord amb els supòsits previs establerts, doncs, d'aquestes obres, *L'Atlàntida* és la que fa més aportacions al lèxic escrit del català contemporani.³

3. En el cas d'algunes unitats lèxiques hi ha dues obres del mateix any que les documenten alhora per primera vegada. Llavors he inclòs la unitat en la llista de cada una de les obres.

Pel que fa al segon grup, si considerem el rànquing per autors — tenint en compte només els autors de dues obres de la llista, que són tres —, el resultat és el següent: Verdaguer (*L'Atlàntida i Canigó*) és l'autor que fa més aportacions, amb 57 unitats, seguit de Narcís Oller (*L'escanyapobres i Croquis del natural*), amb 25, i de Gaietà Vidal i de Valenciano (*La pubilla del mas de Dalt i La família del mas dels Salzes*), amb 23.

Comentaris

Cal tenir en compte que els resultats de l'estudi s'han de considerar indicadors, ja que es va examinar una mostra d'unitats lèxiques del subcorpus (3.500) i no pas la totalitat (77.148) (es va examinar, per tant, un 4,5% del total). Amb tot, la posició de Verdaguer com a autor important pel que fa a la seva aportació al lèxic escrit del català contemporani sembla que queda ben apuntada. D'altra banda, si aquests resultats, segons com, poden semblar contradictoris en relació amb els resultats de l'estudi anterior (el de la riquesa lèxica relativa de la totalitat dels textos del subcorpus), potser fonamenten una de les hipòtesis que suggeria abans: hi pot haver una obra A amb un índex de riquesa lèxica relativa superior a una obra B, però potser el lèxic de l'obra A és un lèxic més comú — més compartit amb el lèxic d'obres anteriors — i el de l'obra B és més original i, per tant, més rellevant. Donaria suport a aquesta idea, per exemple, el fet que, de les obres seleccionades per al segon grup (amb 20.000–30.000 ocurrències), *L'Atlàntida* és, com hem vist, la que fa més aportacions al lèxic escrit del català contemporani; en canvi, obres amb una quantitat d'aportacions menor tenen un índex de riquesa lèxica relativa superior, com és el cas de *Croquis del natural*, *L'escanyapobres* i *Lo prodigi del segle*.

És important en qualsevol cas que recordi que el mètode utilitzat aquí no és l'únic mètode possible per avaluar la rellevància literària del lèxic d'un autor. Quan projectava el treball per avaluar les aportacions lèxiques de Verdaguer a la llengua literària vaig pensar també en una altra via: en comptes de partir d'una mostra aleatòria d'unitats lèxiques (que obliga a dedicar moltes hores a una feina «inútil», com és constatar simplement que una bona pila de les unitats consultades no són protodocumentades en cap de les obres estudiades), podia haver partit d'unitats lèxiques d'o-

bres de Verdaguer. És clar que aquí el risc de trobar que Verdaguer no és el primer autor que documenta les unitats continua existint i, sigui com sigui, si es volia fer un treball comparatiu hauria calgut completar-lo igualment amb l'examen d'unitats lèxiques d'obres d'altres autors.

Una altra possibilitat és canviar de postulats metodològics. Establir, per exemple, que el lèxic escrit del català contemporani no està constituït per les unitats recollides al subcorpus literari del CTILC sinó per les unitats recollides al *Diccionari general de la llengua catalana* (1932), de Pompeu Fabra. I, llavors, examinar en quin autor modern és protodocumentada cada una de les paraules recollides en aquesta obra. I encara es poden fer altres aproximacions, algunes de prou més complexes. Es podria considerar que el fet que un autor protodocumenti un mot no necessàriament vol dir que aquest mot sigui una aportació d'aquest autor al lèxic català contemporani. Si continuéssim amb el supòsit al qual m'acabo de referir que el lèxic escrit del català contemporani és el que recull el *Diccionari general* de Pompeu Fabra, el que seria rellevant és indagar quins són els autors que Fabra va fer servir com a font de cada un dels lemes de l'obra (que potser no són els que primer van fer servir el mot en qüestió). Per exemple, el mot *brandejar* és protodocumentat, al subcorpus literari del CTILC, en l'obra *Celísties*, d'Ubach i Vinyeta. Si busquem el mot al diccionari de Pompeu Fabra, trobem aquest article:

brandejar *v. tr.* Brandar (una arma). *Brandejant amb mà ferrenya l'espasa flamejant.*

L'exemple, com tots els exemples del diccionari de Fabra, és un exemple sense indicació de filiació, però el nostre col·lega Pere Farrés segur que s'apressaria a fer-nos notar que és un vers de *L'Atlàntida*. A més, l'exemple no sols il·lustra el significat, sinó que el determina. A l'obra d'Ubach *brandejar* no significa «brandar una arma». Hi expressa el moviment de les campanes quan repiquen. Fabra, doncs, va seguir Verdaguer en aquest cas, i no pas Ubach i Vinyeta.

Finalment, cal subratllar que, naturalment, les dades quantitatives que resulten de l'estudi i d'altres que caldrà continuar fent en la mateixa línia són susceptibles d'anàlisis de tipus qualitatiu. Per exemple, si comparem els 40 mots «aportats» per *L'Atlàntida* amb els 36 del seu immediat «perseguidor», el poema humorístic *Lo prodigi del segle* del reusenc

Barberà i Canturri, podrem observar que, mentre que entre els mots de *L'Atlàntida* només hi ha un castellanisme (*borronejar*), entre els mots de *Lo prodigi* n'hi ha quatre (*burrical*, *pàrrafo*, *rascada* i *bisarro*).

Final

Al començament de la meua intervenció expressava, en relació amb la possibilitat de mostrar objectivament la riquesa lèxica de Verdaguer i de determinar-ne l'abast, el meu neguit i la meua sensació d'impotència. Després d'haver fet el treball que he presentat en aquestes ratlles, aquests sentiments s'han atenuat. Certament, continuo pensant que és fonamental confeccionar l'inventari lexicogràfic de l'obra de Verdaguer, i que només després d'haver analitzat aquest inventari podrem valorar adequadament l'aportació de l'autor de *L'Atlàntida* al lèxic escrit del català contemporani. Però, alhora, espero haver mostrat que hi ha mètodes d'anàlisi que, mentrestant, permeten aportar dades en la línia que proposava. Mètodes que, a més, també tindran utilitat quan s'hagi elaborat l'inventari complet del lèxic verdaguerià.

Referències

- CTILC = Corpus Textual Informatitzat de la Llengua Catalana. Accessible en línia a través del Portal de la llengua de l'Institut d'Estudis Catalans, <http://pdl.iecat.net> [consultes fetes de l'abril al setembre del 2005].
- FABRA, Pompeu (1932). *Diccionari general de la llengua catalana*. Barcelona: Llibreria Catalònia.
- RAFEL, Joaquim (director) (1998). *Diccionari de freqüències. 2 Llengua literària*. Barcelona: Institut d'Estudis Catalans.
- VERDAGUER, Jacint (1995a [1886]). *Canigó*, edició crítica a cura de Narcís Garolera. Barcelona: Quaderns Crema.
- (1995b [1865]). *Dos màrtirs de ma pàtria, o siga Lluçà i Marcià*, a cura de Ricard Torrents. Vic: Eumo Editorial / Societat Verdaguer.
- (2002 [1877]). *L'Atlàntida*, a cura de Pere Farrés i Arderiu. Vic: Eumo Editorial / Societat Verdaguer.

APÈNDIX

Taula 1. Paraules (de la mostra) que el subcorpus literari del CTILC documenta per primera vegada en les obres de 6.000-7.000 ocurrences que han servit per a l'estudi⁴

Jacint Verdaguer, <i>Dos màrtirs de ma pàtria</i> (1865)	Francesc Ubach i Vinyeta, <i>Celísties</i> (1866)	Tomàs Aguiló i Forteza, <i>Poesies fantàstiques en mallorquí</i> (1852)	Benet Altet, <i>Déu i lo món</i> (1854)
allongar	acoquinar	albó	bull-bull (m)
amazona	arç	colcada	mant (m)
arboç	aurífer	escorpí	
avingut (adj)	boirada	espira	
cruiximent	brandejar		
empiri (m)	caçaire (m)		
envanit (adj)	oriol		
esdevenir (m)			
floresta			
folrar			
girant (m)			
llépol			
obelisc			
onatge			
revellit (adj)			

4. En les taules que presento hi figura, per a algunes paraules que es presten a interpretació equívoca, un codi que n'indica la categoria gramatical: m = nom masculí; f = nom femení; adj = adjectiu; adv = adverbi; interj = interjecció. He seguit l'assignació de categoria gramatical que fa el CTILC, perquè altrament hauria alterat els resultats, que depenen del sistema de lematització específic usat al CTILC. Així, he respectat els participis que el CTILC categoritza com a adjectius, encara que en molts casos una anàlisi més acurada hauria mostrat que els hauria de correspondre la categoria de verbs (i llavors aquests participis deixarien de comptar com a unitats lèxiques independents). En les paraules que comparteixen la primera documentació amb altres obres (que poden ser o no de les seleccionades per a l'estudi), hi he col·locat un asterisc. Dono la forma de les paraules regularitzada, si bé en algun cas hi afegeixo entre claudàtors la forma original que apareix al text.

Taula 2. Paraules (de la mostra) que el subcorpus literari del CTILC documenta per primera vegada en les obres de 20.000–30.000 ocurrences que han servit per a l'estudi (I)

Jacint Verdaguer, <i>L'Atlàntida</i> (1877)	Josep Barberà i Canturri, <i>Lo prodigi del segle</i> (1868)	Gaietà Vidal i Valenciano, <i>La pubilla del mas de Dalt</i> (1866)	Jacint Verdaguer, <i>Canigó</i> (1886)
aixaragallar	abolit (adj)	aixoll	alimara
apetonar	anatematitzar	àpoca	aloja
argentífer	apostrofar	armer (m)	amalecita
arramadar	arrogar	cobdiciar	apocalíptic
arrasador	bisarro	còbit	aufrany
atlàntic	burrical	consums	beduí (m)
àtom	butllofa	contenda	cadellada
boa	casuista	conxa	cèltic
borronejar	corpenta	crac	condigne
brandó	desenterrar	eixugamà	conreria
capitombar	discutit (adj)	espatller	degotís
carnívor	ègloga	forcat (m)	epitalàmic
centaure	enciclopèdia	galejar	estellicó
condolit (adj)*	equivalència	garri (m)	glaçador
congesta	esbombar	llegítima	goja*
coral·lina			
{coralina}	exordi	manual (m)	salobre (adj)
córrec	fanatisme	metòdic	trenat (adj)
dellà [d'allà]	fanatitzar	muntant (m)	
ditada	firmar (interj)	napoleó	
embrancar	horriblement	perdiguer (m)	
emperesir	inimitable	pinso	
endarrerir	innominat (adj)		
endomassat			
(adj)	miner (adj)		
enfurismar	morralles		
espantable	oratória		
fresquívol	orbe*		
fumarel·la	pàrrafo		
grava	pretèrit (adj)		
hamaca	rascada		
joguínos	rebaixa		
liana	reconciliat (adj)		
llenyer (m)*	ressonant		
mandró*	retrocés		
molsós	trabocar		
rèmora 'peix'	trapassar		
rodera	xano-xano (adv)		
sòcol			
solellós			
terboli			
vaquer (adj)			

Taula 3. Paraules (de la mostra) que el subcorpus literari del CTILC documenta per primera vegada en les obres de 20.000–30.000 ocurrences que han servit per a l'estudi (II)

Narcís Oller, <i>L'escanyapobres</i> (1884)	Rosend Arús, <i>Cartes a la dona</i> (1877)	Narcís Oller, <i>Croquis del natural</i> (1879)	Antoni Ferrer i Codina, <i>Lo pagès de l'Empordà</i> (1875)
abaltir	apunt	accessible	ambulància
anorrear	belga (m)	bavejar	bracejar
botjar	borina	conreat (adj)	carquinyoli
brossam	buf (interj)	convalescent (m)	crestat (adj)
campit (adj)	cursi	cupè	espagiat (adj)
codicial	eixancarrat (adj)	desmesura	fai
copsat (adj)	[<i>niño</i>] gordo	figurat (adj)	maniquí
descantellar	gòtic	intransigent	missiva
foradada (f)	ianqui	minuciosament	pàroa
idoni	jipijapa		
forja	mandró*		
guardaagulles	quinze		
mener	samfaina		
reconfortar	sis (m)		
sentenciosament			
solidari			

Taula 4. Paraules (de la mostra) que el subcorpus literari del CTILC documenta per primera vegada en les obres de 20.000–30.000 ocurrences que han servit per a l'estudi (III)

Pilar Maspons i Labrós, <i>Narracions i llegendes</i> (1874)	Francesc Fayos, <i>Plançons</i> (1882)	Josep Martí i Folguera, <i>Versos catalans</i> (1893)	Gaietà Vidal i Valenciano, <i>La família del mas dels Salzes</i> (1880)	Antoni Careta i Vidal, <i>Eures</i> (1882)
batxiller	blasfem	enlluernant	branquejar	collar
brevatge	desmuntar	filtre	solana	reclada
coquesa*	interpel·lació	pirinenc		
enfellonir	provisionalment			
estepa	sufocant			
finalment	transfusió			
reliquiari				
turquesa (f)				