

RESEÑA DE SÁNCHEZ LÓPEZ, E. (2013). ESTUDI DE LA LLENGUA D'AUSIÀS MARCH A TRAVÉS DE LES COL-LOCACIONS. UNA APROXIMACIÓ SEMIAUTOMÀTICA. BERLIN/BOSTON: WALTER DE GRUYTER

El de la lingüística de corpus (LC) per a l'anàlisi literària és encara un terreny poc trillat si el comparem amb altres disciplines on s'ha fet ús d'eines informàtiques (didàctica de la llengua, lexicografia, dialectologia, etc.). I encara més pel que fa a la literatura catalana, terreny que resta erm, a penes, en el sentit com l'autora d'aquest llibre utilitza la LC. No és gens estrany si pensem que, per a molts, l'ús de la lingüística per a l'anàlisi literària és criticable per motius diversos (per a què un analista literari parla de lingüística?, per a què serveixen les xifres i l'estadística si es tracta de literatura?, etcètera). Des del nostre punt de vista sobre literatura i lingüística de corpus, aquesta última ha de ser considerada com una eina més de recolzament per a l'anàlisi literària. Tanmateix, si el seu ús no s'acompanya de la interpretació humana pertinent, no té cap sentit la recerca perquè no és significativa, només serien números i llistes de paraules.

El treball que la professora de la Universitat d'Alacant Elena Sánchez ens presenta en aquest llibre, s'inclou dins d'un projecte col·lectiu molt més ampli, dirigit per Vicent Martines i que respon a l'acrònim IVITRA (Institut Virtual d'Investigació Traductològica). Aquest grup és especialment important per tres motius: un és la publicació incessant de traduccions a llengües diverses, sobretot d'obres clàssiques de la literatura catalana medieval (*Curial i Güelfa*, *Tirant lo Blanch* i poemes d'Ausiàs March, entre d'altres); un altre motiu és la creació d'un corpus amb textos del català antic i modern (CIMTAC), corpus complet i ben etiquetat; el tercer motiu, que pren forma en el llibre que ressenyem, no és altre que el disseny d'una sèrie d'eines de processament i de consulta del corpus que acabem d'esmentar, que permeten aprofitar-lo analíticament a uns nivells més que acceptables en comparació amb altres corpus d'arreu del món.

Pensem que un dels aspectes més innovadors del sistema IVITRA és la possibilitat de recepció de noves formes provinents de nous textos etiquetats per part del diccionari base i gràcies a l'eina Metatagging. És una bona manera de detectar noves formes lèxiques però també supralèxiques provinents del català antic i no sistematitzables fàcilment. De fet, Elena Sánchez, juntament amb Maria Àngels Fuster i altres membres de l'equip (però amb un rol més destacat) han estat les encarregades del disseny d'aquestes eines i de la seua posada en pràctica inicial. De fet, totes dues tenen sengles tesis doctorals on han aprofitat per experimentar amb aquesta infraestructura informàtica (Fuster, 2008; Sánchez, 2010). Sense anar més lluny, aquest llibre és una adaptació de la tesi de doctorat de Sánchez. L'elevat nivell d'etiquetatge que assoleix aquest corpus del català antic i modern és possible gràcies a la prèvia creació de diccionaris de LC a partir de la digitalització i del processament de l'obra d'autors entre els quals destaca Ausiàs March, protagonista de la recerca de l'autora.

En la introducció, Sánchez ens ofereix una retrospectiva molt interessant sobre l'estudi de les col·locacions, sobretot basada en el que anomenarà Escola Britànica, dins la qual trobem Firth (1957), l'iniciador de l'estudi de la dimensió col·locacional de la llengua, i Sinclair (1966), el primer a utilitzar eines informàtiques per a l'estudi de les col·locacions. És apreciable aquesta revisió des dels orígens de l'estudi de les col·locacions, tornant enrere més enllà de la diàspora de treballs sobre el tema des de diferents àmbits (lingüística, didàctica de segones llengües, lexicologia, etc.) que ha tingut lloc en els darrers cinquanta anys.

Per mitjà de les diverses parts del llibre, Sánchez refereix la confusió suscitada per una denominació anàloga de dos conceptes diferents i clau en l'estudi de la fraseologia. El problema rau essencialment en el fet que el que l'escola britànica anomena *collocation* no és el mateix que l'escola espanyola anomena *colocación*. Quant a aquest "malentès terminològic", Sánchez comparteix la posició de Jordi Ginebra, tot defensant que, en català, una *col·locació* és qualsevol coaparició lingüística, mentre que una *concurrència* és quelcom més concret que s'esdevé quan aquesta col·locació o coaparició presenta trets fraseològics (2013: 24).

Com hem dit, l'autora es posiciona a favor dels postulats de Jordi Ginebra, l'introduïdor de les teories fraseològiques de Firth i Sinclair a la tradició catalana, a l'hora de definir els paràmetres que governen la fraseologia, de manera que amb aquesta tendència (i amb d'altres de paral·leles com ara les de Salvador, Piquer i Sancho Cremades) podem intuir una mena d'*escola catalana* com a principi de tradició en l'estudi de la fraseologia de la llengua catalana des d'una perspectiva conciliadora que rep aportacions de tradicions diverses, les digereix i n'aprofita els punts forts.

En aquest sentit, aquesta intuïda *escola catalana* defensaria la importància de mantenir les denominacions anglosaxones pel pes específic d'aquesta tradició, però es mostra receptiva a les aportacions interessants en l'àmbit espanyol, com és el cas dels trets definitoris de les unitats fraseològiques proposats per Gloria Corpas (1996). De fet, Sánchez basa la seua descripció de les característiques de les unitats fraseològiques en l'obra de Corpas (p. 28). A més, s'hi introduiran les escoles fraseològiques alemanya, pel que fa a la lexicogramàtica, l'ensenyament de la llengua estrangera i el concepte de verbs suport (p. 150), i francesa (p. 32).

Així, entenem que la introducció d'aquest llibre és clara i completa. Aporta un desglossament detallat de les diferents edicions de l'obra marquiàna i ho fa des de l'interessant punt de vista del multilingüisme, punt de vista influït pel doble vessant de professora i de traductora de l'autora. Sánchez destaca la presència de glosses i d'edicions multilingües des de les primeres edicions de March, i aquestes glosses ja incloïen unitats polilèxiques. Tanmateix, li ha calgut cercar un aparat teòric consistent. En paraules de l'autora: "atés que els glossaris no inclouen cap reflexió teòrica i que les investigacions de Sinclair i Halliday se centren bàsicament en la detecció i selecció mitjançant el criteri únic de la freqüència, hem hagut de recórrer al marc teòric d'una disciplina que ens permetera l'anàlisi i classificació de les unitats polilèxiques trobades. A fi de cobrir aquesta necessitat hem recorregut al marc teòric de la fraseologia" (p. 27).

D'altra banda, és destacable com Sánchez soluciona el problema sempre afegit a l'hora de considerar les col·locacions en l'obra literària d'un autor. Ho fa considerant dues dimensions: les col·locacions en general de la llengua i les col·locacions de l'estil de l'autor (pp. 41-42). Pensem, però, que aquestes dues dimensions en generen una tercera en el seu inevitable punt de col·lisió: les col·locacions que ho són en la llengua general i en l'idiolècte literari. Malgrat que, com hem dit adés, Sánchez s'esforça per deixar ben definides col·locacions i concurrències, el concepte de *cluster* no s'esmenta, de manera que sovint hi ha dubte de si quan es refereix a concurrències parla de combinacions consecutives o bé amb altres mots intercalats i canvis d'ordre.

L'autora no tractarà la qüestió de la lingüística de corpus, que tant subratllem en aquesta ressenya, fins entrats en l'apartat de metodologia i anàlisi. Com hi mostra Sánchez, el procés d'etiquetatge genera mots dubtes, que en el seu cas s'han intentat resoldre sempre pensant en futures aplicacions més enllà de la seua recerca (és a dir, més enllà del català medieval, de cap dialecte i de l'idiolècte d'un autor). Com que li interessen les col·locacions, Sánchez opta per cercar els mots més forts semànticament, sobre els quals s'articularen probablement les col·locacions fraseològiques. És per això que cerca substantius, verbs, adjectius i adverbis.

Més endavant, l'estudi dels *bigrames* i *trigrames* que presenta Sánchez (pp. 55-63) mostra com, tot i les precaucions que es prenguen al respecte, és impossible evitar el soroll informatiu en els resultats de les cerques (es pot reduir, però mai eliminar). El pes inevitable de la intuïció del filòleg es demostra, per exemple, a l'hora de triar les categories lèxiques a cercar en els *bigrames* (ex. adj. + N) i els *trigrames* (ex. prep. + adj. + N). Cal dir que, en aquesta part del llibre, les anàlisis de les col·locacions són de caire semàntic i s'evita la citació d'exemples concrets i la contextualització o explicació pragmàtica dels casos. El següent punt de l'anàlisi, que és el més extens, es basa en la combinació de dues cerques que permeten la creació de diverses graelles de coaparicions. D'una banda, tenim les cerques dels 40 substantius i els 40 verbs (se cerquen aparicions literals però apareixen en les graelles de forma lematitzada) i la selecció de *bigrames* i *trigrames* triats per l'autora i organitzats per categories lèxiques. D'aquesta combinació se'n derivaran una sèrie de col·locacions que l'autora sistematitzarà en la primera de les dues parts analítiques d'aquest llibre.

La freqüència, generalment considerada un dels trets definitoris de les unitats fraseològiques, perd la seua importància quan apliquem la LC a l'anàlisi literària. Aquest handicap té relació també amb el problema de la manca d'accés a un possible corpus de control (si no se'l fabrica cada investigador) que té la LC aplicada a la literatura, i que s'agreuja en el cas de la literatura catalana (p. 70): o bé te'l fas tu (com és el cas del CIMTAC) o no n'hi ha.¹ L'únic corpus de control possible el trobarem quan analitzem l'estil d'un poemari en concret i el corpus de control serà tota l'obra de l'autor en qüestió. Sembla que aquest problema el tenen també en altres llengües, fins i tot l'anglès, quan es tracta de LC i literatura.²

¹ Existeixen alguns intents en aquest sentit, com el corpus de l'Institut d'Estudis Catalans, però encara no permeten cerques prou refinades a l'usuari extern.

² Pensem, per exemple, en l'estudi que Bettina Fischer-Starcke presenta en el seu llibre *Corpus Linguistics in Literary Analysis: Jane Austen and her Contemporaries* (2010).

D'altra banda, no sembla que pel fet de ser el gènere poètic el dels textos del corpus, la forma dels versos afecte o influeixi massa o gens en l'anàlisi (no trobaríem, així, cap diferència significativa amb l'anàlisi de prosa literària). Només es nota perquè l'autora té en compte les conseqüències de l'exacerbació de la funció poètica en la deformació de les col·locacions de la llengua, en la forma del text (com ara la irregularitat sintàctica per quadrar el metre i la rima, que es tracta en la pàgina 152 del llibre).

Es copsa una acurada reflexió sobre la metodologia per part de l'autora, que afirma que "totes les eines, informàtiques o conceptuals emprades en la consecució del present estudi persegueixen un objectiu principal: dur a terme una anàlisi del conjunt complet –per això la referència a la descripció global, en un sentit d'extensió topogràfica" (p. 154).

Sánchez tanca amb la reflexió sobre la següent pregunta: "on rau la 'literarietat' de March?" Definitivament, cal transcendir el nivell lèxic i anar al supralèxic per a respondre aquesta pregunta. La principal conclusió de l'autora sobre l'estil i literarietat del poeta és que March innova en les combinacions de mots més que no en els mots pròpiament, ja que s'hi detecten pocs neologismes i poca varietat temàtica a partir del lèxic. En aquest sentit, l'autora destaca la "creativitat col·locacional" del poeta (p. 155). Quelcom semblant passa amb el poeta Vicent Andrés Estellés.

Arribats al zenit de la segona dècada del segle XXI, quan llegim textos sobre la *moderníssima* lingüística de corpus escrits al començament dels noranta (com ara Sinclair, 1991 dins Sánchez, 2013: 44) se'ns fa estrany. Potser és ja hora de deixar de referir-se a conceptes de lingüística de corpus tradicionals (etiquetatge, concordances, programes com el Word Smith Tools, *clusters*, diccionaris de LC, bases de dades) com quelcom pioner o avantguardista. Llàstima que no tinguem encara un bon Banc del Català (anàleg al corpus Bank of English) etiquetat i fàcilment consultable. Fins on hem pogut comprovar, els intents de corpus del català realitzats fins ara es caracteritzen per ser parcials, sovint privats o bé molt difícils de consultar i amb escasses possibilitats en les cerques.

La lingüística de corpus aplicada als estudis literaris crea amors i odis: des dels obsessionats de les llistes als quals els arbres no els deixen veure el bosc fins a l'extrem contrari, habitat per filòlegs o humanistes al·lèrgics a les xifres estadístiques. El nostre raonament a propòsit és tan antic com Aristòtil: la justa mesura és la mare de totes les coses (o hauria de ser-ho). Així, com déiem al començament, defensem l'ús de la lingüística de corpus al servei de l'anàlisi literària, no mai desconnectada del context cultural més ampli, que és "el bosc" en la metàfora que encara ara ressona.

Això sí: si s'utilitza la lingüística de corpus per a l'anàlisi literària, o en qualsevol altre camp, ha de ser de manera justificada i científica, que és com ho ha fet Elena Sánchez. En aquesta recerca doctoral convertida en llibre subjau la idea de la lingüística de corpus per a l'anàlisi literària com a nou punt de vista, més sistemàtic, transversal i multifuncional, que pot servir per refermar, rebatre o bé matisar prèvies anàlisis sobre els mateixos autors i obres fets sense l'ajuda dels prismàtics de la lingüística de corpus. Tot comptat i debatut, pensem que l'estudi que Sánchez ofereix en aquest llibre està molt ben dissenyat i sistematitzat, té coherència global i les hipòtesis i la metodologia emprada són ajustades: el llibre és suficient per mostrar una visió completa i alhora aprofundida de la recerca que presenta.