

I Simposi Internacional de Lexicografia
Institut Universitari de Lingüística Aplicada
Barcelona, 16-18 de maig de 2002

La lexicografia i la identificació automatitzada de neologia lèxica

M. T. Cabré, M. Domènech, R. Estopà, J. Freixa, E. Solé
Observatori de Neologia (IULA)

1. El concepte de neologisme

El significat del mot neologisme és difícil de delimitar amb paràmetres objectivables com reflecteixen els diccionaris generals de llengua en definir-lo. Segons el *Diccionari de la Llengua Catalana* un neologisme és una "unitat lèxica nova, formalment o semànticament, creada en una llengua per les pròpies regles de formació de mots o manlevada a una altra llengua". El *Diccionario de la lengua Española* (RAE, 1992) ens ofereix una definició molt semblant: "vocablo, acepción o giro nuevo en una lengua". Le Robert (1993) s'hi refereix com "emploi d'un mot nouveau ou emploi d'une expression préexistants dans un sens nouveau. Mot nouveau; sens nouveau d'un mot". El Coubuil () "neologism is a new word or expression in a language, or a familiar word or expression taht is now being used with a new meaning; a formal word". Totes aquestes definicions remetent al concepte de nou "originat o ocorregut fa poc, aparegut de poc, que apareix per primera vegada", concepte altra vegada difícil d'objectivar perquè ofereix uns paràmetres molt relatius per saber quines unitats lèxiques de la llengua es poden considerar noves i ens suggereixen interrogants com els següents que no tenen una resposta unànime: què vol dir exactament "fa poc"? fa poc per a qui? fa poc respecte de què? com sabem que és la "primera vegada" que un mot apareix? apareix on?

Diversos autors han remarcat la relativitat del concepte neologisme i, tradicionalment, han recorregut als criteris que els parlants tenim per identificar les paraules noves per definir aquest concepte. A, Rey, per exemple, considera que una unitat lèxica és neològica segons quin sigui el paràmetre d'identificació del qual es parteix. Aquest

autor estableix tres paràmetres de diferent naturalesa per identificar un neologisme que ens permeten establir conjunt de mots nous diferents: temporal, psicolingüístic i lexicogràfic. Així, si el criteri de selecció és temporal, considera neologisme qualsevol paraula que ha aparegut en un període recent; si el paràmetre és psicolingüístic, un neologisme serà aquella unitat lèxica que un parlant (o un col·lectiu de parlants) percep com a nova; i, finalment, si el criteri és lexicogràfic, considera que una unitat lèxica es pot considerar neològica si no es troba documentada en un determinat corpus lexicogràfic. (GILBERT) (EL DE LIMOGES, ELISABET)

Tots aquests criteris d'identificació presenten inconvenients i avantatges, i si els analitzem ens adonem que els podem minimitzar a dos: criteri monovariàble i criteri plurivariàble. El primer és objectivable i sistematitzable i, en canvi, el segon és més difícil d'objectivar i delimitar.

El criteri monovariàble també es pot anomenar documental en el sentit que ens permet considerar neològiques paraules que no es troben documentades en un corpus, ja sigui lexicogràfic o textual, prèviament establert.

El criteri plurivariàble no depèn només de si hi són o no en un corpus de referència, sinó que responen a diversos factors lligats a la percepció de novetat que un parlant té respecte d'una unitat de la llengua. Aquest criteri és molt més subjectiu perquè està condicionat per paràmetres de diferent naturalesa com el context comunicatiu, el cabal de lèxic, el coneixement enciclopèdic, sociocultural, el coneixement lingüístic, de cada individu. És doncs un criteri poc sistemàtic que dificulta poder establir uns paràmetres objectius de detecció de neologia.

2. El concepte de neologisme de l'Observatori de Neologia

L'Observatori de Neologia és un grup de recerca de l'IULA de la UPF que ja compta amb dotze anys de treball en aquest àmbit està format per un grup molt ampli i diversificat de col·laboradors tant pel que fa a la seva formació com a la seva activitat professional. Aquesta diversitat dels seus membres ha condicionat des dels orígens la selecció del criteri de buidatge de neologismes. Així, per superar la subjectivitat alhora de detectar i de decidir si una paraula és nova o no l'Observatori de Neologia s'ha centrat en el criteri lexicogràfic.

--Corpus Què buidem? neologismes dels mitjans de comunicació d'àmplia difusió en una primera els de premsa escrita i després audiovisuals i de textos escrits espontanis (és a dir, no filtrats per serveis de correcció).

-- establiment del criteri lexicogràfic, fonts de referència

--utilització en el buidatge manual, certa limitació també és una mica subjectiu pel que fa al coneixement de les fonts de referència

--utilització en el buidatge automàtic, molt útil D'altra banda, i a mesura que s'han anat diversificant les noves tecnologies, un segon objectiu prioritari de l'Observatori de Neologia ha estat i és aconseguir avançar en les vies d'automatització del procés de buidatge de neologismes; és a dir, en la identificació i selecció de les unitats, i en la compleció del banc de dades, tenint en compte les diferents característiques de les fonts de buidatge. els corpus lexicogràfics són l'element clau en la cadena de detecció, selecció i buidatge de neologismes; però també mostrarem com els corpus lexicogràfics són (o poden ser) receptors o punt d'arribada del resultat neològic obtingut al final de l'automatització d'aquest procés. Però com ja hem esmentat anteriorment, el criteri lexicogràfic també té inconvenients en la identificació automatitzada de la neologia lèxica: en primer lloc, perquè resulta insuficient, ara per ara, per detectar determinats tipus de neologismes (per exemple, la neologia semàntica, és a dir, noves accepcions d'entrades ja documentades; i dins de la neologia formal, la neologia formada per sintagmació). Caldrà trobar mecanismes per avançar, en la mesura del possible, en la detecció d'aquestes unitats si no de forma automatitzada, sí de forma assistida.

--utilitat del criteri lexicogràfic en relació amb les aplicacions, l'explotació: actualització dels diccionaris. L'anàlisi d'aquests neologismes permet fer el seguiment de la vitalitat lingüística del català i del castellà, llengües amb què es treballa, i servir de font per a l'actualització de diccionaris de llengua general (prescriptius i descriptius) o per a l'elaboració de diccionaris descriptius de paraules noves.

-- El criteri lexicogràfic ens ha estat molt durant aquesta primera dècada de l'Observatori tant en el buidatge manual de la neologia com sobretot en el buidatge automàtic. Com en les aplicacions.

Ara bé, després d'una dècada i de tenir una base de dades de neologismes molt gran, si s'analitza, ens adonem que dins de la base qualsevol parlant de la llengua s'adona que hi ha molt soroll, és a dir que hi ha moltes paraules que tenint en compte el criteri intuïtiu, altrament dit psicolingüístic, no serien neològiques. i no es tracta pròpiament d'un inconvenient, utilitzar com a filtre de detecció les obres lexicogràfiques aconduïx cap a l'obtenció d'unitats amb diferents graus de neologicitat: (a) unitats no documentades en un determinat corpus lexicogràfic, que poden estar incloses en altres fonts; (b) determinats *neologismes de llengua* en la terminologia de Guilbert (1975), que no acostumen a formar part de la nomenclatura dels diccionaris bàsicament per raons de predictibilitat; (c) alguns neologismes consolidats en el banc de dades (per tant, poc neològics ja des d'un punt de vista temporal i psicolingüístic), que no han estat incorporats en els diccionaris per raons sociolingüístiques i pragmàtiques.

Si s'analitza amb detall podem observar que els neologismes es poden discriminar, filtrar en funció del grau de neologicitat. Es poden establir graus diferents de neologicitat que respondrien a criteris ...

Per tant, arribem a la conclusió que com ja intuïen alguns autors el concepte de neologismes no serà mai ni únic ni sistematitzable ni objectivable a partir només d'un criteri. D'aquesta manera, ens serveix el concepte de prototipus proposat per la lingüística cognitiva ja que permet establir uns neologismes que des de molts punts de vista són els més prototípics...i que segurament respondria al conjunt d'unitats que la majoria de parlants d'una mateixa societat sociocultural considerarien noves de manera que la resta d'unitats no considerades prototípiques s'anirien allunyant progressivament d'aquests exemplars...fins "aire de família"

