

**Joaquim Rafel**

Membre de la Secció Filològica

i director del projecte Diccionari del Català Contemporani

## **Preàmbul**

És un fet conegut que la lexicografia ha experimentat en els darrers decennis una pregona i intensa renovació metodològica. A partir de l'inici dels anys seixanta sorgeix la consciència que s'acaba un llarg període de la història de la lexicografia i que s'inicia una nova època del treball lexicogràfic, que s'ha de caracteritzar per una tasca cada vegada més sistemàtica i objectiva, per una descripció de la llengua cada vegada més acurada i per la incorporació dels ordinadors com a elements integrants d'aquesta activitat. Aquesta nova època s'ha anat obrint pas, primer lentament i després d'una manera cada vegada més ràpida; els primers articles crítics sobre aspectes concrets del mètode lexicogràfic i les primeres obres de conjunt han estat seguits per una munió de títols, tant sobre aspectes particulars com sobre qüestions de caràcter general, que s'han multiplicat amb profusió; en aquests anys, a més, han sorgit una sèrie de publicacions especialitzades en lèxic i en lexicografia. La bibliografia sobre aquestes qüestions és avui gairebé aclaparadora.

Aquest procés renovador sorgeix com a reacció davant el caràcter asistemàtic de la lexicografia tradicional, caracteritzada fonamentalment per l'acumulació de dades, sovint d'una manera acrítica, per l'absència d'unes bases teòriques i d'un mètode explícit i —salvat algunes obres comptades— per la servitud respecte de les obres lexicogràfiques anteriors. A l'origen d'aquest moviment renovador trobem, en canvi, la necessitat d'incorporar a l'activitat lexicogràfica, d'una banda, els avenços que s'han produït en el camp de la lingüística en els darrers decennis i, d'una altra banda, les innovacions tecnològiques que afavoreixen el tractament massiu de la informació, amb l'aparició i el ràpid avenç de la informàtica.

Ara no podem analitzar detalladament cada un dels aspectes afectats per aquesta renovació; només esmentaré que una de les fites principals de la nova lexicografia és l'objectivació del mètode: la subjectivitat i la intuïció deixen d'ocupar un lloc preeminent en l'activitat dels lexicògrafs i són substituïdes per les dades objectives. I en aquest procés d'objectivació del mètode els corpus textuals informatitzats tenen un paper molt destacat. La utilització d'un cor-

pus textual adequadament representatiu com a font principal en el procés d'elaboració d'un diccionari es converteix així en una de les característiques més rellevants de la lexicografia actual, fins al punt que s'estableix una frontera clara entre els diccionaris basats en corpus i els diccionaris no basats en corpus.

Voldria destacar des d'ara que aquest canvi d'orientació metodològica influeix d'una manera determinant en diversos aspectes de la configuració d'un diccionari; caldria precisar potser també que quan parlem de la utilització de corpus textuals com a base de l'elaboració dels diccionaris no ens referim a una utilització de la informació que conté el corpus com a mer auxiliar o complement d'una tasca lexicogràfica duta a terme a partir d'uns altres supòsits metodològics, a un ús accessori, com quan es restaura un edifici vell i se li dona una aparença de nou, cosa també possible, sinó que al·ludim a la utilització, des del moment de la concepció del diccionari, de les dades fornides pel corpus, de tal manera que aquestes dades en condicionen l'estructura i la configuració tant en aspectes molt generals com en aspectes més particulars.

## **El català i la lexicografia basada en corpus**

És en aquest context metodològic que ja fa uns quants anys, el 1984, la Secció Filològica de l'Institut d'Estudis Catalans prengué la decisió d'iniciar els treballs que havien de conduir a l'elaboració d'un diccionari descriptiu de la llengua catalana dels darrers cent cinquanta anys, és a dir, aproximadament des de la Renaixença, com a base d'un futur diccionari normatiu; aquesta decisió anava lligada a la idea que per a prendre de la manera més adequada les decisions que afecten la normativa lingüística cal disposar d'una descripció tan exhaustiva com sigui possible de la llengua amb criteris científics i sobre unes bases empíriques, cosa que lamentablement no sempre és possible. És per això que, com a primera fase o fase prèvia d'aquest projecte, fou concebuda l'elaboració del Corpus Textual Informatitzat de la Llengua Catalana (CTILC), un corpus suficientment extens i representatiu per a poder servir de base a l'elaboració del dit diccionari. La constitució del corpus s'ha dut a terme entre 1985 i 1997, i actualment està totalment acabat i etiquetat morfosintàcticament i disposa dels mecanismes de recuperació de la informació adequats per a la tasca lexicogràfica. Tot i que el seu objectiu principal és la utilització amb finalitats lexicogràfiques, a l'hora de constituir-lo es prengueren les mesures necessàries perquè tingués un caràcter multifuncional, i és així que avui, a més de la utilització específica en lexicografia, pot ser usat per a qualsevol tipus de treball lingüístic que requereixi l'anàlisi de dades textuals.

No és ara el moment de detallar les característiques d'aquest corpus; només diré que el CTILC està constituït per un total de més de tres mil textos, publicats entre 1832 i 1988, i que l'extensió total supera els cinquanta milions de mots de text; aquests es distribueixen en un 44 % corresponents a llengua literària i un 56 % corresponents a llengua no literària. Els textos, d'altra banda, han estat classificats tipològicament, cosa que permet disposar de la distribució de les

diferents unitats del corpus entre els diversos tipus de text; per a la llengua literària s'han classificat d'acord amb els gèneres tradicionals i per a la llengua no literària s'ha optat per un mètode mixt que combina els diferents temes o camps de coneixement i certs tipus funcionals de text.

En aplicació del programa de treball establert des de l'inici, després de l'acabament del corpus, durant aquests dos darrers anys (1998 i 1999) s'han establert les bases metodològiques i s'ha fixat l'estructura del nou diccionari descriptiu que s'ha d'elaborar sobre la base del corpus i s'han aprovat els criteris que han de presidir-ne la redacció. Després de fer les proves de redacció que s'han considerat necessàries, en els darrers mesos s'ha iniciat la redacció sistemàtica dels articles.

## Corpus i diccionari

Sembla evident que a l'hora d'elaborar un diccionari descriptiu de la llengua basat en un corpus, són només les dades ofertes pel corpus les que s'han de tenir en compte en la redacció de l'obra, i encara potser prescindint d'aquells mots o d'aquells significats utilitzats per un sol autor en una sola obra per a donar solució a una necessitat expressiva d'una ocasió particular.

A l'hora d'elaborar un diccionari normatiu, tot i que hom haurà de tenir en compte factors diversos, i fins i tot diferents canals d'alimentació, d'acord amb els criteris d'establiment de la normativa lingüística que s'apliquin, crec que es pot afirmar que el fet de disposar d'un corpus suficientment extens i representatiu ofereix una base de discussió nova sobre els mots i els usos que hi han de figurar i els que no hi han de figurar; aquest fet pot contribuir, indubtablement, a alleugerir el pes mort que per a tot diccionari representen els mots o les accepcions de presència no justificada, i, per una altra banda, pot augmentar-ne l'operativitat en permetre la incorporació de mots i sentits d'ús viu, justificat de manera empírica, absents en els diccionaris elaborats prèviament.

## Macroestructura

Si fixem la nostra atenció de moment només en els aspectes macroestructurals, és a dir, en l'inventari de mots que conté un diccionari i els que sorgeixen de l'anàlisi d'un corpus, trobem ja diferències significatives. L'existència en català del Corpus Textual Informatitzat de la Llengua Catalana ens permet de comparar la situació en els diccionaris amb la d'un corpus suficientment extens i equilibrat per a poder ser considerat representatiu de la llengua del període cronològic que comprèn. Com que disposem també en format electrònic de la nomenclatura dels diccionaris que constitueixen el punt de referència lexicogràfic fins a la data dels darrers textos que formen part del corpus (*Diccionari general de la llengua catalana*, de Pompeu Fabra, amb les addicions de l'Institut d'Estudis Catalans, i *Diccionari de la llengua catalana*, d'Enciclopèdia Catalana, 2a

ed., 1983), podem establir una comparació rigorosa i sistemàtica entre el corpus i aquests diccionaris. El resultat d'aquesta comparació ofereix unes dades numèriques més que sorprenents; el corpus té un total de 149.075 lemes, i el conjunt d'aquests diccionaris en té 87.994. Per a obtenir la diferència entre un i altre inventari, no podem simplement restar aquestes xifres, perquè hi ha solament un cert nombre de lemes que són compartits pels dos conjunts: exactament 57.913; hi ha, per tant, 91.162 lemes del corpus que no es troben en els diccionaris i 30.081 lemes dels diccionaris que no són presents en el corpus, és a dir, el 61 % dels lemes del corpus no tenen presència als diccionaris i el 34 % dels lemes dels diccionaris no són representats al corpus. Tot i les matisacions que podríem fer a aquestes xifres si poguéssim entrar ara en una anàlisi no merament quantitativa sinó qualitativa (criteris de lematització, mots de freqüència baixa que corresponen a singularitats expressives o a allò que s'anomena *mots d'autor*, mots realment usats que no apareixen en els diccionaris per criteris normatius, etc.), de tota manera el resultat d'aquesta comparació continuaria sent espectacular. A la vista d'aquesta espectacularitat, no puc deixar de pensar en les opinions controvertides que s'han expressat sobre el valor dels diccionaris i el valor dels corpus textuals com a representació de la llengua. Al costat d'opinions que defensen que el fet de ser enregistrat en un diccionari qualsevol consagra l'existència d'un mot com a element lexical de la llengua, o que un inventari fonamentat en textos no té més probabilitats de representar el lèxic real de la llengua que un repertori lexicogràfic, en trobem d'altres d'absolutament contràries, com que la pertinença d'un mot a una llengua no es mesura per la seva inserció en un diccionari; fins i tot, dins aquesta segona línia, el fet d'utilitzar materials procedents dels diccionaris per a la realització d'estudis sobre el lèxic ha estat acusat de donar una imatge poc fidedigna de la llengua. No entrarem ara en la discussió d'aquestes opinions tan contraposades, però no les podem perdre de vista en qualsevol de les consideracions que fem sobre aquest tema.

Si fem una aproximació més de caràcter qualitatiu a les dades generals que acabem de presentar, observant la situació en alguns grups de mots formats amb procediments més o menys previsibles (prefixació o sufixació), podem constatar que es produeix una situació comparable a la que hem observat en la totalitat de les dades. Només en posaré un exemple: si prenem en consideració els mots formats amb el prefix *bio-*, observarem que en el corpus hi ha 133 mots amb aquest prefix i en els diccionaris de referència n'hi ha 99; d'aquests dos conjunts, només hi ha 42 mots que coincideixen, els del corpus i els del diccionari; això vol dir que n'hi ha 91 que apareixen en el corpus i no apareixen en els diccionaris i 57 que apareixen en els diccionaris i no en el corpus. Entre els que són exclusius del corpus, trobem mots com *bioclimàtic* (f = 3), *biocombustible* (f = 1), *biodinàmic* (f = 7), *bioelèctric* (f = 4), *bioestadística* (f = 3), *bioindicador* (f = 1), *bioinsecticida* (f = 4), *biomèdic* (f = 6), *biomatemàtic* (f = 4), *biomolecular* (f = 1), *biopolímer* (f = 4), *biopsíquic* (f = 5), *biosensor* (f = 5), *biotecnologia* (f = 196), *biotecnològic* (f = 52), *biotò* (f = 2), etc.; entre els que apareixen en els diccionaris de referència i no en el corpus tenim mots com *bioclima*, *bioconversió*, *biodeterioració*, *biodinamisme*, *bioflavonoide*, *biofotogènesi*, *bioherm*, *biòmetra*, *biometria*, *bioreacció*, *bioritme*, etc.; i entre els que apareixen tant en

els diccionaris de referència com en el corpus, tenim, entre altres, *biodegradable* (f = 13), *biofísica* (f = 8), *biogènesi* (f = 9), *biòleg* (f = 159), *biologia* (f = 480), *biològic* (f = 1.280), *biomassa* (f = 106), *biòpsia* (f = 21), *bioquímic* (f = 70), *bioquímica* (f = 53), *biosfera* (f = 104), *biosíntesi* (f = 14), *biòtop* (f = 65), etc.

Si ens fixem, d'altra banda, en mots que presenten una relació sistemàtica previsible, com són, per exemple, els que es formen amb un prefix negatiu, trobem també nombroses diferències entre el corpus i els diccionaris, i, en aquest cas, no és estrany de trobar que les dades freqüencials que ens ofereix el corpus es decanten en favor del mot absent en els repertoris lexicogràfics; posaré també només algun exemple d'aquest fet: el mot *deformable*, que apareix en els diccionaris, presenta en el corpus una freqüència menor que el mot *indeformable*, absent en els repertoris lèxics; el mateix passa amb la parella *repetible* i *irrepetible* i amb d'altres. D'altra banda, trobem també contradiccions entre mots d'estructura anàloga: *incontenible*, que apareix en el corpus amb una freqüència digna de ser presa en consideració, no apareix en els repertoris lèxics, i, en canvi, sí que hi apareix *incongelable*, que en el corpus apareix una sola vegada. Casos com aquests que he esmentat a manera d'exemple són molt nombrosos. Deixeu-me dir de passada que algunes d'aquestes incongruències ja han estat reconduïdes al *Diccionari de la llengua catalana* de l'Institut, que inclou alguns d'aquests mots com *irrepetible* i *incontenible*, però no, en canvi, *indeformable*, tot i mantenir *deformable*.

Pel que fa a aquests mots que formen part de les possibilitats creatives del lèxic a través de procediments sistemàtics, podríem considerar que el diccionari, a través de la inclusió dels mots identificats en la realitat observable del discurs, ens hauria de donar la informació verídica sobre els límits factuais de l'aplicació de les regles que en permeten la formació, però els fets ens mostren que en el diccionari hi ha una sèrie de mots d'utilització dubtosa en la llengua real, i, contràriament, molts dels mots utilitzats en el llenguatge corrent, presents fins i tot en els textos escrits, no són recollits en els diccionaris.

A partir de la constatació de fets com aquests, podríem arribar a deduir que la inclusió o no de determinats mots derivats en el diccionari es guia per criteris basats en virtualitats més que no pas en realitats. És com si diguéssim: «Incloem *repetible*, però no *irrepetible*, que és merament un derivat previsible d'aquest amb sentit negatiu, l'existència del qual l'usuari pot deduir a partir d'unes regles conegudes», sense tenir així en compte que les necessitats comunicatives fan més probable la utilització del mot *irrepetible* que no pas la del mot *repetible*.

Com ja he dit fa un moment, a la base d'aquestes opcions hi ha opinions contraposades sobre la prioritat de criteris sistemàtics sense un suport empíric comprovat o de criteris basats en l'ús real de la llengua; no cal dir que deixem de banda el rerefons ideològic que hi pot haver a la base d'aquestes opinions contraposades.

Aquests aspectes que acabem de comentar afecten l'estructura de la nomenclatura d'un diccionari, però no tenen una transcendència normativa; ningú no interpretarà que infringeix la norma lingüística pel fet d'utilitzar el mot *indeformable*, encara que no sigui al diccionari normatiu. En altres casos, però, la informació que proporciona el corpus enfront de la que conté el

diccionari pot tenir una transcendència d'aquest ordre. En posaré un parell d'exemples que afecten la forma dels mots: en el diccionari normatiu tenim els mots *conhortar* i *consolar*, però no *aconhortar* ni *aconsolar*; en el corpus trobem tant els uns com els altres, però amb valors freqüencials contraposats; en el cas de *consolar*, aquesta forma és la més freqüent ( $f = 779$ ), enfront de *aconsolar*, que té una freqüència més baixa ( $f = 386$ ), considerable, tot amb tot; però en el cas de *conhortar*, aquesta forma és la menys freqüent ( $f = 92$ ), enfront de *aconhortar*, que té una freqüència més elevada ( $f = 252$ ); en el primer cas, la relació freqüencial va a favor de l'opció del diccionari, però en el segon les dades afavoreixen clarament l'opció contrària. Això convida a la reflexió, encara més si tenim en compte que en un altre cas comparable (*contentar* enfront de *aconcentrar*) l'opció del diccionari normatiu reconeix la forma *aconcentrar*, que és també la més freqüent en el corpus ( $f = 1024$ ), i exclou *contentar*, que és la menys freqüent en el corpus ( $f = 345$ ). Uns altres exemples relatius a la forma dels mots poden ser els de *temorós* i *temerós*, més problemàtics; el primer, l'únic que reconeix el diccionari normatiu, presenta una freqüència més baixa ( $f = 186$ ) que el segon, que el diccionari exclou ( $f = 353$ ), o el dels mots *tesar* i *tensar*; en aquest darrer cas ens trobem que *tensar* és més freqüent que *tesar* en el corpus, i no apareixia en els repertoris lexicogràfics de referència que hem esmentat, però ara ha estat incorporat al diccionari normatiu, al costat de *tesar*.

Deixant de banda els problemes de forma, en el corpus hi ha, a més, molts mots que no apareixen en els repertoris de referència; molts d'aquests són castellanismes evidents o mots tinguts per castellanismes, com, per exemple, *plàtica* ( $f = 110$ ), *sublevació* ( $f = 229$ ), *puro* ( $f = 195$ ), *gasto* ( $f = 1023$ ), *desterro* ( $f = 137$ ), *despedir* ( $f = 894$ ), *quantios* ( $f = 136$ ), *certer* ( $f = 71$ ), *dàdiva* ( $f = 26$ ), *morador* ( $f = 59$ ), *modals* ( $f = 46$ ), *viril* (part de la custòdia) ( $f = 18$ ). Altres, en canvi, són mots previsibles, almenys formalment, d'acord amb les regles de formació de mots del català (sigui per prefixació o per sufixació, per composició o per conversió sintàctica), no recollits, però, en els diccionaris de referència, com, per exemple, *absolut* (subst. masc.) ( $f = 164$ ), *angoixant* (adj.) ( $f = 42$ ), *autodomini* (subst. masc.) ( $f = 14$ ), *carada* (subst. fem.) ( $f = 10$ ), *culturalment* (adv.) ( $f = 129$ ), *vertaderament* (adv.) ( $f = 181$ ), *disjuntiva* (subst. fem.) ( $f = 38$ ), *inconcret* (adj.) ( $f = 84$ ), *lleument* (adv.) ( $f = 284$ ), *obertament* (adv.) ( $f = 413$ ). Aquests que acabo d'esmentar no han estat inclosos per ara en el diccionari normatiu; altres, en canvi, que tampoc no figuraven en els repertoris lexicogràfics que ens han servit de punt de referència, ja figuren en el diccionari de l'Institut, com, per exemple, *barracó* (subst. masc.) ( $f = 24$ ), *bussejar* (verb) ( $f = 12$ ), *comissionat* (subst. masc.) ( $f = 170$ ), *commiseratiu* (adj.) ( $f = 13$ ), *decorat* (subst. masc.) ( $f = 491$ ), *delictiu* (adj.) ( $f = 59$ ), *inigualable* (adj.) ( $f = 27$ ), *nerviosisme* (subst. masc.) ( $f = 123$ ), *rampinya* (subst. fem.) ( $f = 45$ ), *sext* (adj.) ( $f = 104$ ), *tardoral* (adj.) ( $f = 169$ ), *voluntariós* (adj.) ( $f = 67$ ), i altres.

Deixem, però, de banda les qüestions relacionades amb la macroestructura del diccionari, que —malgrat la seva importància— no constitueixen sinó una part molt petita de la contribució dels corpus a la configuració d'un diccionari, per a poder dir uns mots en relació amb els aspectes microestructurals.

## Microestructura

Des d'aquest punt de vista, podem dir que la utilització d'un corpus textual com a base per al disseny i l'elaboració d'un diccionari determina molts aspectes de la naturalesa dels articles, tant des del punt de vista del contingut, com de l'organització interna d'aquest contingut; i això no és sinó una conseqüència lògica del fet que el lexicògraf, en el moment de redactar un article determinat, disposa d'una informació objectiva i exhaustiva sobre la unitat lèxica que l'encapçala.

Pel que fa a la determinació dels significats de la unitat lèxica, hom pot partir de l'anàlisi dels diferents contextos en què apareix, i arribar a establir la xarxa de significacions, en clar contrast amb els diccionaris no basats en corpus, que han de recórrer necessàriament a la tradició lexicogràfica anterior, a la intuïció del lexicògraf, o, en el millor dels casos, al buidatge selectiu de textos, amb la càrrega de subjectivitat que això suposa.

La informació freqüencial —inherent al concepte de *corpus*—, a part del valor merament informatiu, que indubtablement té, pot convertir-se en un element operatiu a l'hora d'estructurar el diccionari. La utilització més transcendent d'aquesta informació és la de classificar els diferents significats d'un element lèxic, és a dir, les diferents accepcions d'un article, amb criteri freqüencial; o, si més no, fer intervenir aquest criteri, al costat d'altres, en l'organització jeràrquica dels significats.

Una altra de les aportacions destacades de la utilització d'un corpus com a base del treball lexicogràfic és la informació exhaustiva que podem obtenir sobre les estructures sintàctiques de què pot formar part una unitat lèxica, i sobre les coocurrències lèxiques més freqüents; aquest és un tipus d'informació que manca gairebé del tot en els diccionaris convencionals, en els quals, si de cas, s'ha de cercar d'una manera indirecta, i sol estar en general molt lluny de donar un reflex fidel de la realitat de l'ús, sigui perquè és insuficient o asistemàtica, sigui perquè és inexacta o inadequada, o, simplement, perquè no hi és.

Altres qüestions relacionades amb la microestructura tenen una certa complexitat, i és per això que no ens podem entretenir a il·lustrar-les totes amb detall, perquè això requeriria un temps i un espai de què no disposem. Posaré només un exemple senzill d'alguna de les possibles conseqüències de la utilització d'un corpus en la informació que conté un article. L'examen del material del corpus relatiu al mot *calor*, en el sentit de '(sensació produïda per una) elevada temperatura ambiental', ens mostra que aquest mot es combina habitualment amb un nombre reduït de verbs, amb presència o absència controlada de l'article, i dóna lloc a algunes expressions que, per bé que molt usades, no són d'interpretació òbvia: *fer calor*, *tenir calor*, *passar calor*, *suportar la calor*, *entrar en calor*; d'altra banda, els adjectius que es combinen habitualment amb aquest substantiu són de caràcter magnificador (*forta*, *intensa*, *extraordinària*, *excessiva*), o bé indiquen qualitats negatives o no desitjables (*abrasadora*, *aclaparadora*, *asfixiant*, *enganxosa*, *espantosa*, *horrible*, *inclement*, *inexorable*, *insuportable*, *sufocant*, *terrible*, entre altres), sense excloure la combinació amb termes neutres, en menor escala (*estival*, *humida*, *natural*, *primavera*, *seca*, *solar*); hi ha l'expressió *fa una calor que empudega*, per a referir-se a una calor molt

intensa i insuportable, i la frase *no em (et, li...) fa ni fred ni calor*, en el sentit de *tant me (te, li...) fa*, molt habituals. Doncs bé, encara que pugui sorprendre, gairebé cap d'aquestes informacions no apareix en cap dels diccionaris generals a què pot recórrer l'usuari, ni en el *Diccionari català-valencià-balear*, tot i la pretensió descriptiva, històrica i dialectal que té; el diccionari Fabra duu *tenir, sentir calor* com a exemples, sense cap explicació, cosa que recull el diccionari de l'Institut; els diccionaris d'Enciclopèdia Catalana, en canvi, suprimeixen aquests exemples i incorporen l'expressió *fer calor*, que expliquen.

Aquestes consideracions, preses aïlladament ja tenen, em sembla, la seva importància; però, si les projectem en un marc lexicosemàntic més ampli, aquesta importància creix. Si prenem en consideració el mot *escalfor*, que des del punt de vista del significat conceptual té unes connexions molt estretes amb el mot *calor*, resulta que, a diferència d'aquest darrer, es combina d'una manera gairebé exclusiva amb adjectius que denoten qualitats desitjables o positives (*agradable, agradosa, amable, amiga, bona, benefactora, confortable, dolça, grata, tendra, sanitàosa, retornadora*, i altres); d'altra banda, la modificació apreciativa, que és magnificadora o pejorativa per a *calor* (*calorota, calorassa*), és minimitzadora o meliorativa per a *escalfor* (*escalforeta*).

A qualsevol observador ingenu li pot semblar que els fets lingüístics que acabem de destacar en relació amb aquests mots són pertinents per a l'elaboració d'un diccionari; això no obstant, cap lexicògraf no els ha fet objecte d'atenció preferent a l'hora de redactar els diccionaris catalans de què disposem fins ara.

Des del punt de vista de les possibles repercussions d'un mètode com aquest en un diccionari normatiu, és fàcil d'imaginar que una informació detallada i exhaustiva sobre aspectes com a) la xarxa significativa completa d'una unitat lèxica, avalada per l'ús, b) la relació exhaustiva de les estructures sintàctiques en què participa el mot, i molts d'altres, reconeguts o no per la normativa actual, permeten disposar d'un marc de reflexió sobre molts i molts aspectes que en un futur, esperem que no gaire llunyà, estarem en condicions de discutir sobre la base d'un coneixement empíric de la realitat de la llengua.

Els criteris de redacció del diccionari descriptiu que hem començat a elaborar preveuen que aquest diccionari recollirà totes les formes, tots els usos sintàctics, tots els significats que es dedueixen del corpus, tant si s'ajusten als criteris normatius actuals, com si no s'hi ajusten, distingint, però, amb una codificació adequada aquells que recull la normativa actual d'aquells que no recull, tant si els rebutja explícitament, com si no hi fa referència. No cal que faci remarcar a un auditori com aquest el valor d'una informació com la que resulta de l'aplicació d'aquest mètode per a un propòsit com el que ara ens ocupa.

## Conclusions

Pel que fa a conclusions particulars, crec que si ens prenem moltes de les consideracions que acabo de fer com una invitació a la reflexió, el millor és que després d'haver reflexionat, cada u



en tregui les seves, que poden ser molt diferents, segons el tarannà propi, les conviccions arrelades, l'obertura intel·lectual, els horitzons professionals, o, fins i tot, com he suggerit al començ, segons els seus plantejaments ideològics sobre el concepte de llengua.

Per la meua part, em limitaré a unes consideracions que intentaré que siguin màximament objectives, lligades a l'enunciat general de la meua intervenció, *Corpus i lexicografia*, a les quals, com a generals que són, podríem fer molts matisacions:

1) La naturalesa, l'estructura i la fesomia d'un diccionari dissenyat i elaborat a partir de les dades que ens forneix un corpus textual representatiu de la llengua que es vol descriure poden donar a aquest una imatge o un aspecte molt diferents d'aquells que ens ofereix un diccionari redactat seguint els mètodes convencionals de la lexicografia tradicional.

2) Un diccionari elaborat a partir d'aquests plantejaments ofereix a l'usuari unes garanties que no pot oferir un diccionari tradicional, perquè *a)* tots els mots que conté la nomenclatura, a part del valor que els pot conferir el fet de pertànyer a un sistema lingüístic, tenen el valor particular i concret que els atorga el testimoni de l'ús, *b)* les diferents significacions de cada unitat lèxica tenen una base empírica absent en els diccionaris convencionals, i *c)* tant els mots com els significats poden ser valorats a partir de la freqüència en el corpus, cosa que dóna una dimensió nova a la informació continguda en el diccionari.

3) En resum, la garantia que ofereix un diccionari tradicional convencional és una garantia indirecta —els altres diccionaris—, subjectiva i limitada —el coneixement que els lexicògrafs tenen de l'estructura i funcionament de la llengua—, mentre que la garantia d'un diccionari basat en un corpus és una garantia directa, objectiva i general —el mateix corpus.