

RECENSIONS

MARTIN, Robert (2001): *Sémantique et automate. L'apport du dictionnaire informatisé*. Paris: Presses Universitaires de France, 192 p.

Introduction (11-5)

Fondements et finalités d'une sémantique automatisable (17-66)

I. Les fondements théoriques (17-50)

II. Les finalités et les moyens: le dictionnaire informatisé (50-66)

La réduction polysémique (67-135)

I. L'appariement textuel (68-90)

II. Le filtrage du sens (91-101)

III. Aménagement du dictionnaire et techniques de traitement (101-35)

Compréhension automatique et moteur d'inférence à fondement lexical (137-80)

I. Compréhension automatique et calcul inférentiel (138-54)

II. Vers un moteur d'inférence à fondement lexical (154-80)

Conclusion (181-3)

1. Vet aquí un llibret magistral, d'allò més entretingut pels aspectes que toca, i excellent pels que suscita. No és estrany, perquè es tracta d'una síntesi curta i entenedora —però en cap cas trivial— sobre la intersecció entre la semàntica i l'automatització, dues dimensions temàtiques de gran abast i actualitat entre lingüistes de tota mena (sobretot semantistes, lexicògrafs i especialistes en computació). També ho és perquè està escrit per una autoritat en plena maduresa que sens dubte s'ha divertit força escrivint aquest text que només ell podia compondre fent tanta exhibició de domini temàtic, de claredat expositiva i de luxe clarificador en les exemplificacions.¹ Tot això fa que el llibre sigui una eina immillorable per a introduir-se en les matèries que toca. Inevitablement, a vegades resulta una mica massa resumit, i fins i tot lacònic; p. ex. puntualment, en la presentació oposicional que introdueix entre 'polisèmia d'accepció' i 'polisèmia de sentit' (32) o més en general quan fa cites pròpies, que quasi sempre dona per prou, potser massa, ben conegudes (143, entre altres).

Una altra important qualitat derivada és que lliga la teoria lingüística, i fins i tot lògica, amb la tecnologia computacional (a remarcar la diferència, essencial, que fa des de l'inici entre teoria i tecnologia) i, per tant, basteix un pont que cal aprofitar des del punt de vista pedagògic, atès el divorci, per no dir la ignorància, que hi ha actualment entre molts lingüistes *à la page* i una tradició que, si fos prou coneguda, estalviaria certs descobriments de la mediterrània i encara alguna niciesa i tot. Gairebé emociona veure Bröndal, Hjelmslev o fins i tot Bréal enmig de tot el debat teòric. Omet, en canvi, alguna referència, que trobo inevitable, a la hipòtesi de Sapir-Whorf (23)

De pas, el llibre representa, per ell mateix i per l'entorn conceptual en què es mou, encara una prova de la vitalitat de la lingüística francesa, però també sobre la llengua francesa i en llengua fran-

1. Entre d'altres, exceheixen, al meu parer, les de *curieux* (94) i *bout* (108).

cesa, per si de cas cal recordar-ho, enfront de l'actual monopoli de, en, sobre, per a, des de... l'anglès. Per a ser més precisos, el *Trésor de la langue française* (d'ara endavant *TLF*) —accessible a <http://atilf.inalf.fr/tlfv3.htm> (a setembre de 2002), millor que no pas a l'adreça del llibre, i de consulta imprescindible— es converteix en el punt de referència a l'entorn del qual gira tota l'obra.

2. Per raons de claredat, decideixo orientar la present recensió partint d'una declaració de principis amb una enumeració sumària d'alguns problemes generals que presenta la lexicografia i veure la solució o la resposta que en dona el llibre. Consta que es tracta d'una mera estratègia expositiva i no de crear un marc teòric per examinar l'autor. Ell en sap molt més que jo, de tot això! Com és sabut, l'obra lexicogràfica per excel·lència és el diccionari, una eina que recentment ha evolucionat en forma molt considerable, sobretot amb la incorporació de registres provinents de corpus de referència textuals, és a dir, de reculls, generalment immensos, de textos escrits i orals de la llengua en qüestió. Tot i així, gairebé sempre els diccionaris convencionals encara continuen essent, bàsicament, una col·lecció alfabèticament ordenada d'entrades —amb unes característiques massa conegudes perquè ara les analitzem—. L'important és que aquestes llistes de mots aïllats, sigui quina sigui la informació que duguin, no poden formar part per si soles d'un model integral de la gramàtica. La connexió entre les accepcions dels mots i el seu ús en texts reals depèn quasi sempre de la intuïció lingüística del parlant... sobretot nadiu. Per això no es pot esperar que la consulta de l'entrada *de* aclareixi, a ningú que ho ignori, el sentit d'una construcció com ara *si de cas*, o qualsevol altra. També ho demostra el fet extern que, des de l'escola fins a les facultats filològiques, una cosa és el llibre de gramàtica i una altra el diccionari —ambdós convencionals, clar—. El cas és que el diccionari tampoc no forma part del coneixement intuïtiu de la llengua, sinó, com a molt, d'una reflexió sobre la llengua. Quan un nen de tres anys pregunta a sa mare què significa un mot desconegut que acaba de sentir fa una autèntica reflexió de lingüista, aïlla el mot i el converteix en objecte propi d'atenció. En condicions normals, el mot en qüestió adquirirà per a ell una constel·lació de sentits possibles nebulosament diferenciats que la seva intuïció anirà copsant sense esforç, fins i tot en usos metafòrics, metonímics, irònics i en jocs de mots. Això ens duu al següent corollari: la connexió entre llengua i parla, en el vell sentit de Saussure, o entre virtualitat i realització, no és assumible per cap diccionari convencional, i no per manca de capacitat descriptiva, sinó, senzillament, perquè la virtualitat no té límits (però sí pautes, o regles, Chomsky *dixit*). Per molts milions de referències que incorporin els corpus textuals, i per molta automatització que s'hi posi, la virtualitat no queda encerclada si el sentit dels mots queda inevitablement reduït a un nombre fix d'accepcions. Les referències —cap multitud de referències— no són el sentit de cap mot.² Tornem ara al llibre.

3. Si n'hagués de resumir tot d'una el contingut, diria que Robert Martin, i l'equip que treballa en això, es proposa convertir el *TLF* en un diccionari que superi aquestes mancances, almenys en el sentit —gens modest tanmateix— de fer-lo capaç d'interpretar el contingut de textos reals: «nous nous intéresserons exclusivement à l'analyse» (51). Una tasca extraordinària que marca sens dubte una fita en la història de la lexicografia i, per descomptat, també de la lingüística. És clar que es tracta d'una tasca irrealitzable, això per començar, però, es podria trobar un repte més engrescador?

Com hem vist a la presentació, l'obra es distribueix en tres parts. La primera, dividida en dues subparts, presenta els problemes bàsics de la relació entre semàntica i lexicografia o, altrament dit, de com n'és de fiable i completa la representació dels significats als diccionaris i, per altra banda, classifica tres tipus reconeixibles de diccionaris:

2. Tot i que no distingeix 'sentit' d'altres possibilitats, com ara 'denotació', l'autor estableix constantment la diferència radical entre sentit i referència: «confondre le *sens* et la *référence*, c'est rendre insoluble la question de la polysémie» (40), afirma. Això sense referir-nos encara a l'empobriment de recursos que representa la llengua escrita enfront de la llengua parlada, on els parlants exhibeixen tots els recursos de les seves intuïcions.

- els convencionals, que tothom coneix
- els informatitzats, accessibles a l'ordinador amb opcions d'hipertext, entre altres, i
- els automatitzats, aptes, a més a més, per analitzar textos reals i extreure'n el contingut

Precisament, la segona part es dedica a l'exposició de les tècniques lexicogràfiques pròpies dels diccionaris informatitzats i la tercera fa el mateix amb les tècniques pròpies del diccionari automatitzats, sempre a partir del *TLF* —que queda, així, convertit en diccionari automatitzat o, millor, en vies d'automatització.³

Per tot això comença enumerant, presentant i rebatent tot un seguit, prou complet, de posicions més o menys fonamentades que en la tradició lingüística i filosòfica s'han oposat de molt diverses maneres al supòsit segons el qual el significat —en el llibre gairebé sempre referit com a «sens», 'sentit'— dels mots i dels textos és inaprehensible o, si més no, irrepresentable. L'exposició, generalment sumària, sol mostrar el ferm convenciment de l'autor de les seves conviccions, apuntalades no sols amb arguments, sinó també amb il·lustracions força efectives. No tindria sentit que m'entretingués ara en els detalls, però sobre aquesta part voldria fer algun comentari general, sobretot per la transcendència que té sobre la concepció general de l'obra, encara que no per al seu desenvolupament. Primer de tot, a favor de la majoria dels arguments que l'autor addueix (amb una visió resumida de les principals actituds semàntiques de la història), tenint sobretot en compte que és ben conscient dels problemes entre difícils i insolubles que es poden plantejar. De fet, no s'amaga d'admetre que sense validar, encara que sigui a tort i a dret, la certesa en el raonament i els procediments lexicogràfics, el llibre... no seria possible. La informàtica és, per descomptat, el gran recurs. Però tot admetent que «le dictionnaire informatisé peut accroître sensiblement les performances d'un analyseur sémantique» (51), tampoc no perd de vista que, almenys el *TLF*, «n'est évidemment pas conçu dans une visée d'automatisation. S'adressant à un consultant humain, il fait [...] une large place à l'implicite» (68). Indiscutable.

Un cop acceptada aquesta fonamental justificació metodològica, podem recular una mica en la teoria i posar en dubte l'afirmació —ambigua, si més no— que fa en dir que «dans la connaissance des langues, les dictionnaires jouent à l'évidence un rôle de premier plan» (28). Com he dit, res no autoritza a creure que en el cervell del parlant hi ha un diccionari ni poc ni molt semblant a cap diccionari existent, convencional, informatitzat o automatitzat. Ni en la ment del francoparlant més qualificat hi ha el *TLF*, ni cap d'encara més modèlic, com aquell pel qual malden els lexicògrafs de l'Institut National de la Langue Française en la tasca immensa, i magnífica, que s'han posat pel davant. Com també he dit, els diccionaris convencionals ni tan sols són cap component integral de la gramàtica.

Per descomptat, la informàtica té els seus avantatges, fins i tot decisius, enfront del processament humà. Així ho reconeix l'autor quan declara, per exemple (81), que la integració de moltes paraules en locucions més o menys fixes representa un obstacle sovint importantíssim perquè els humans aprenguin llengües, mentre es converteix en un ajut de primer ordre per al tractament informatitzat de les col·locacions i altres construccions complexes, i per a obrir la porta de la combinatòria —la sintaxi— a la lexicografia. Obrir la porta de la sintaxi a la lexicografia és precisament el que permet integrar el lèxic en un model gramatical d'ús real, i convertir la lexicografia en una eina integrada en la gramàtica. Llàstima que l'objectiu és inassolible més enllà de les expressions més o menys fixes. Almenys des d'un diccionari com el *TLF*, renuncia de penetrar a fons en els as-

3. Abans de seguir, voldria apuntar que hi ha encara un altre tipus de diccionari, que s'utilitza en lingüística computacional o informàtica (o encara, com en diuen els especialistes, en el processament del llenguatge natural) i que forma part integral d'un programari capaç de dur a terme, en conjunt, tasques d'anàlisi, generació o traducció de textos en diversos nivells de complexitat. Aquests diccionaris *ad hoc*, que, per no augmentar la terminologia, podríem anomenar 'totalment automatitzats' presenten una fesomia que no té res a veure amb cap dels altres, ja que les entrades venen definides, no per accepcions, sinó per matrius de trets sintactico-semàntics.

sumptes sintàctics: «comme [les procédures] relèvent de la syntaxe, et non pas de la sémantique, nous ne les approfondirons pas ici: l'analyse syntaxique est en soi un immense domaine de recherche que [...] nous supposons en bonne voie de résolution» (80). Bé, si no fos perquè la sintaxi només té resoltes les qüestions més bàsiques.⁴

Dues consideracions s'imposen. La primera és que Martin no ignora ni defuig cap dels problemes insolubles que hem citat. Al contrari, no deixa de reparar en els usos metafòrics (48, 84, 86), tròpics (sinecdòquics, metonímics..., 30 i altres) i de doble sentit en general, tot i que més aviat negligeix els irònics o litòtics i fins i tot els creatius (com l'adjectiu que acabo de fer d'esma a partir de 'lítote'), i moltes de les inferències derivades del coneixement del món: «certes, la limite entre connaissances lexicales et connaissances du monde est loin d'être précise» (41). I reconeix perfectament que el lèxic de les llengües naturals es comporta d'acord amb una immensa quantitat de fenòmens que van, sense cap frontera definida, dels molt generals (com ara l'ampliació polisèmica) als molt concrets (p. ex., les ironies), de manera que els primers accepten fins i tot una formalització mentre que els segons formen un calaix de sastre, una tremenda casuística (19-20): «laissez à des règles universelles la bride sur le cou, et vous verrez apparaître des singuliers effets» (36). Aquesta observació, per cert, també forma part dels rudiments de la lingüística computacional.

La segona consideració és que, per molt restringit que sembli l'àmbit d'aplicació de la metodologia que exposa Martin, la quantitat de teoria que comporta és immensa, sorprenentment immensa. Seguim, doncs.

4. Com he dit, la mare del problema es troba en la impossibilitat de conjuminar llengua i parla o, en aquest cas, de descriure la virtualitat polisèmica, contínua o analògica, dels significats reals amb una llista fixa, discreta o digitalitzada, d'accepcions. Com quadricular una nebulosa. Martin n'és ben conscient: «la polysémie lexicographique se fonde sur un modèle discret: les acceptions se présentent comme mutuellement exclusives. Or la matière sémantique a toutes les apparences d'un continuum» (42). Admet «qu'il y ait une part d'arbitraire dans les découpages opérés, c'est incontestable» (50). Encara que hi ha punts discutibles, «on optera donc [...] pour un modèle à la fois construit (même si toute la construction est en soi labile) et discret (même si, par nature, le sens est un objet fluide aux configurations indéfiniment variables)» (50) i apeña a la multiplicació d'accepcions com a única sortida —en el símil, condensant la nebulosa en tantes gotetes com pot—. Trobo, però, que simplifica en excés quan afirma que «pourtant les faits de continuité ne rendent pas caduque, au contraire [sic], la représentation discrète» (47). Digui, com a molt, que no hi ha cap altra sortida a la vista.⁵ Ara bé, és una sortida prou acceptable, certament, si el nombre d'accepcions és prou gran —com la música digitalitzada que ofereix una impressió contínua de tots els matisos registrables per l'oïda.

Un cop traspassat aquest llindar teòric, l'exhibició de recursos de Martin comporta, al meu parer, la part més atractiva del llibre. Voldria presentar, si més no concisament, alguns aspectes de la magistral estratègia metodològica d'anàlisi semàntica que ens duu fins als límits d'explotació atesos en el *TLF*. Per començar, estableix quatre nivells de càlcul semàntic:

— predicatiu, per tal de reduir l'ambigüitat polisèmica. És el punt de partença, on s'estableix el conjunt d'accepcions d' $A = \{x, y, z, \dots\}$

— referencial, per identificar les entitats al·ludides i crear un àmbit espaciotemporal, el conjunt de coses a què pot al·ludir, o referir-se, A . És el testimoni de la parla, on «les problèmes posés

4. Una cosa semblant passa en l'explotació morfològica. En conjunt es pot aprofitar bona part de la composició derivativa i molta de la flexiva, però en canvi és pràcticament impossible establir connexions automàtiques entre vulgarismes i cultismes (cf. *nit* i *nocturn*) i en general mots etimològicament relacionats (*rei* i *regla*), així com evitar falses atribucions (*rata* i *ratificar*).

5. Com s'explica, si no perquè es copien, que els diccionaris (convencionals) coincideixin tant en l'estructura, en les definicions i sovint fins i tot amb les exemplificacions?

sont d'une difficulté considérable» (54). De fet, «une difficulté quasi insurmontable tient au fait que la référence [...] s'opère indifféremment, par les mêmes marqueurs, ou bien au monde ou bien au texte» (55)

— modal, per a establir el tipus de certesa amb què s'expressen les proposicions, i

— textual, per a extreure una interpretació global (52). És la reunió de totes les dimensions, que comporta no sols l'entrada a la sintaxi, sinó, nogensmenys, al coneixement del món (58...), una gosadia que obre fronts sovint només assequibles a la intuïció dels parlants: «tout cela crée pour l'analyse automatique des pièges redoutables. Une chose est sûre: les connaissances linguistiques n'y suffisent pas. Des *basses de connaissances* sont indispensables» (59). Per això quan es tracta de passar de l'accepció del mot a la d'un text, i es pregunta «qu'est-ce le sens d'un texte? Il serait bien présomptueux de vouloir répondre à une telle question. On dira seulement que comprendre un texte [...] c'est être capable d'en extraire l'essentiel et de le réorganiser dans un abrégé synthétique; c'est [...] le traduire sous forme de *résumé*» (60-1). Déu n'hi do. Més endavant reconeixerà que «une séquence de mots peut signifier plus que l'addition de ses parties» (85). No: sempre és així.

Per a dur a terme tot això distribueix el contingut de la continuació en dues parts estratègiques. Una estratègia de primer nivell consisteix a obtenir una reducció polisèmica (67-135) mitjançant dos procediments: l'aparellament textual (68-90) i el filtratge (91-101), amb un conjunt de subprocediments i cinc regles (74-7), que afronten tanmateix una considerable casuística (79...). Aplica criteris sintàctics, de construcció i combinatoris, i semàntics per al filtratge, de subcategorització, classes lèxiques, camps semàntics i dominis (91). Després de comprovar de nou la gran casuística i l'agudesia requerida per a la interpretació dels exemples, conclou: «quel automate entrera jamais dans de telles subtilités? Mais de tels exemples ne doivent tout de même pas décourager. La proportion des cas traitables devrait être considérable» (101). Ara bé, imposant diversos 'arranjaments' de dades en forma d'esforços de sistematització (102-4), d'explicitació (104-7) i de complementació (107) —que aplica a una sèrie d'exemples il·lustratius (108-21) sobre alguns textos (121-34)—, obté un percentatge considerable d'èxit, a l'entorn del 90%. La resta queda a cura de l'expert o de l'usuari: «quant aux 10 % résiduels, ils peuvent soit être traités par l'expert, si l'on considère que le système doit fonctionner sous la forme d'un 'traitement assisté par ordinateur', soit, si la finalité est celle d'une automatisation intégrale (mais inévitablement problématique), faire l'objet d'une approche plus syntétique». Confesso que m'hauria agradat que hagués estat més explícit aquí, perquè, quan acaba afegint «mais c'est-là une autre manière encore de réaménager les données...» (135), es fa inevitable pensar en els models de gramàtiques semàntiques de la lingüística computacional i preguntar-se si hi ha cap relació.

L'estratègia de segon nivell, que arriba a la màxima ambició metodològica, consta de la comprensió automàtica i el motor d'inferència «à fondement lexical» —important, però lògica, precisió!— (137-80):

— La comprensió automàtica (138-54) comporta un càlcul inferencial segons que la comprensió s'entengui com que 1) «l'automate comprend une phrase *p* s'il est capable de donner de *p* une représentation dans un système déterminé, [...] la traduction de *p*, à partir du langage où *p* est exprimé, dans un langage de représentation» (138), és a dir, que sigui assumible, p. ex., per una base de coneixement; o bé com que 2) «l'automate comprend une phrase s'il est capable de tirer les conséquences sémantiques que cette phrase détermine» (= l'ensemble de phrases vraies si *p* est vraie, l'ensemble de ses inférences) (138). De nou compareix una gran, irreductible, casuística entre les inferències sintètiques i analítiques (140-7). A remarcar l'esquema de 146 per a tots els tipus d'inferència assenyalats. Admet, a més a més, dues variants de 'compréhension': a) «faible: l'automate est seulement en état de dire, à partir de *p* et de *q*, si *q* est ou n'est pas une conséquence de *p*, i b) forte: l'automate est en état, à partir de *p*, de produire lui même la conséquence *q* (139).

— Al seu torn, l'elaboració d'un motor d'inferència de base lèxica (154-80) requereix 1) «un dictionnaire qui explicite les formes propositionnelles et les inférences que ses définitions portent en elles»; i 2) «un système qui, moyennant ce dictionnaire, engendre un traitement approprié» (168). En aquesta part, la més ambiciosa, recorre a un càlcul temàtic que defineix com una «procédure simplifiée, calcul inférentiel affaibli» (153) Recorda de nou les gramàtiques semàntiques.

Com construir, en definitiva, un motor inferencial? En el cas del *TLF* no hi pot haver dubte, si no és a partir de les definicions: «les définitions jouent là un rôle déterminante. Par les formes propositionnelles que l'on peut y associer et les inférences qu'on en tire, les définitions constituent potentiellement un système d'une formidable richesse. Mais pour mettre en place un tel système, il faut préalablement distinguer toutes les composantes définitoires et, dans toute la mesure du possible, apprendre à l'automate à les reconnaître» (154). Això comporta un treball ben intens sobre l'elaboració de definicions, començant per una classificació en tipus: hiperonímica (156-7), sinonímica (157-8), derivativa/translativa (158-9), mereonímica (part, 'mereònim' / tot 'holònim') (159), aproximativa (159-60) i metalingüística (160). Falta, al meu parer, l'exemplificativa, que típicament es dona en les entrades de colors (*blanc* [del color de la neu], *vermell* [del color de la sang arterial, de les roselles]...).

Un cop reconeguda encara una altra important casuística farcida d'excepcions (164-80), ve la culminació, la síntesi: «ces traitements sont-ils automatisables? Le passage de la définition aux inférences analytiques de formes propositionnelles demande de toute évidence l'intervention de l'expert —et le travail à réaliser est considérable» (178) I acaba dient: «reconnaissons que la part automatisable [...] est relativement faible. Il faut convenir aussi que le dictionnaire "automatisé" s'écarte plus encore qu'on pouvait le penser du dictionnaire initial, même "informatisé". Mais celui-ci est un point de départ indispensable. Son réaménagement est le prix à payer pour rendre opératoire l'immense richesse qui recèle» (180).

A la conclusió tanmateix, tot i venir just a continuació, l'autor repren una actitud més aviat programàtica: «le dictionnaire informatisé, peu à peu aménagé en dictionnaire automatisé, constitue une pièce maîtresse d'un système destiné à la fois à la levée des ambiguïtés polysémiques et à la production des inférences nécessaires à la compréhension automatique» (182). Però acaba fent honor a l'evidència: «mais gardons-nous avant toute chose de nous payer d'illusions! N'oublions pas que les quelques orientations ici esquissées ne portent que sur un seul des modules du système (le "module prédicatif"): rien n'est dit du calcul référentiel, du calcul modal, du calcul textuel, où le dictionnaire n'est pratiquement d'aucun recours» (182).

Com totes les obres ben fetes, aquesta desperta una multitud de qüestions que van molt més enllà del seu propi abast. A part del que s'hi aprèn, per descomptat.

Ramon CERDÀ
Universitat de Barcelona

GOHARD-RADENKOVIC, Aline (1999) : *Communiquer en langue étrangère. De compétences culturelles vers des compétences linguistiques*. Bern, Berlin, Bruxelles, Frankfurt/M., New York, Wien: Peter Lang. 253 p.

En première partie, Aline Gohard-Radenkovic dresse un «état des lieux, enjeux, représentations et conceptions» de la didactique du Français Langue Étrangère [17-98]. Elle évoque la profonde mutation que subit l'enseignement des langues étrangères ces dernières années.

Auparavant, l'apprentissage des langues étrangères représentait, certes, une matière dans la formation d'un élève, mais peu étaient ceux qui, par la suite, dans le cours de leur vie professionnelle,