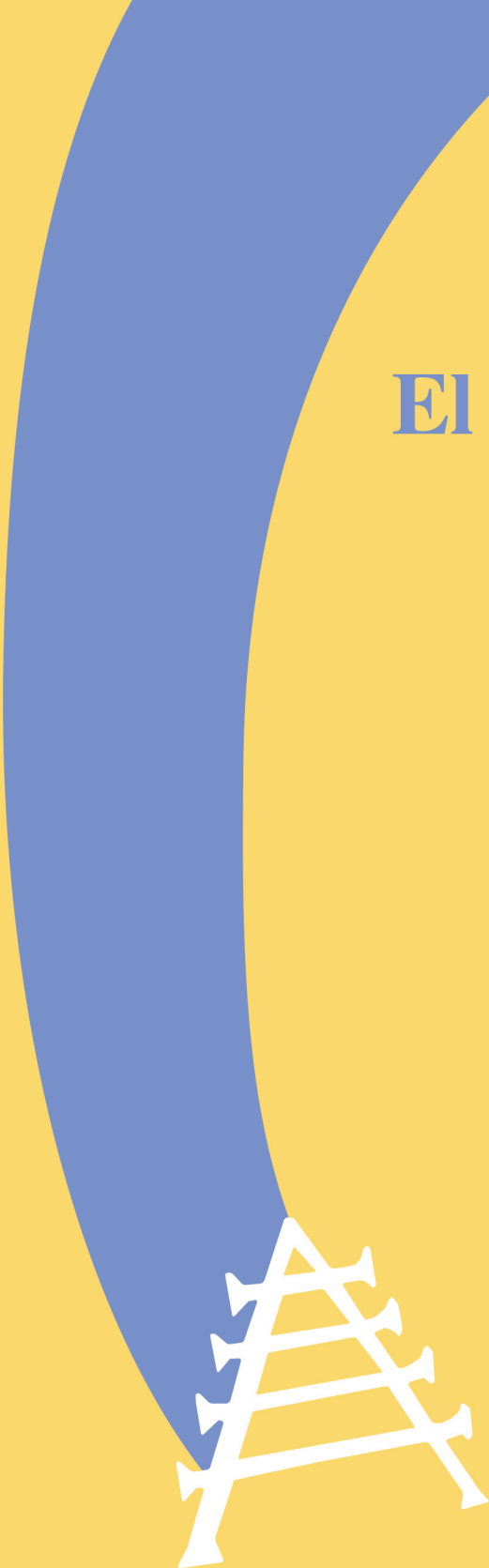




El canvi lingüístic

Teories i eines computacionals per estudiar-lo

Àngels Massip-Bonet
Maria-Pilar Perea (eds.)

UBe

El canvi lingüístic

El canvi lingüístic

Teories i eines computacionals per estudiar-lo

Àngels Massip-Bonet
Maria-Pilar Perea (eds.)



UNIVERSITAT DE
BARCELONA

Edicions

© Edicions de la Universitat de Barcelona

Adolf Florensa, s/n

08028 Barcelona

Tel.: 934 035 430

comercial.edicions@ub.edu

www.edicions.ub.edu



© dels textos respectius: Vicent Salvador, Àngels Massip-Bonet, Irene Castellón Masalles, Gemma Bel-Enguix, Maria-Pilar Perea

Organització del 28è Col·loqui Lingüístic de la Universitat de Barcelona (CLUB 28): Àngels Massip-Bonet, Maria-Pilar Perea. El Grup d'Estudi de la Variació (GEV) i el projecte «Scripta i variació dialectal» (FFI2016-80482-P) de la Universitat de Barcelona han contribuït a la publicació d'aquest llibre.

Becaris: Roman Alonso i Fernàndez, Òscar Casado

ISBN: 978-84-1050-056-3

Data d'edició: 2023

Aquest document està subjecte a la llicència de Reconeixement-NoComercial-SenseObraDerivada de Creative Commons, el text de la qual està disponible a: <http://creativecommons.org/licenses/by-nc-nd/4.0/>.



Índex

Introducció <i>Àngels Massip-Bonet, Maria-Pilar Perea</i>	9
Toc toc —qui és?—. El canvi (reflexions pragmadiscursives) <i>Vicent Salvador</i>	13
Variació lingüística i canvi: aproximació des de la perspectiva dels sistemes complexos adaptatius <i>Àngels Massip-Bonet</i>	27
Mètodes de processament per a l'anàlisi lingüística de la variació <i>Irene Castellón Masalles</i>	45
Cap a una quantificació de l'evolució semàntica del lèxic mitjançant tècniques basades en corpus diacrònics i word2vec <i>Gemma Bel-Enguix</i>	65
Modelitzacions matemàtiques, aprenentatge automàtic i ciència ciutadana per a l'estudi del canvi (socio)lingüístic <i>Ernest Querol Puig, Emília López-Iñesta</i>	83
Canvi lingüístic i anivellament dialectal <i>Maria-Pilar Perea</i>	103

Introducció

El 28è Col·loqui Lingüístic de la Universitat de Barcelona va tractar del canvi lingüístic des d'una perspectiva que pretenia integrar aspectes teòrics i aplicacions concretes.

Reunim aquí un conjunt de capítols —que en la seva primera versió van constituir les conferències del 28è CLUB de la UB— en els quals a través de la variació que es presenta en textos o en enquestes o en altres reculls de dades i dels canvis que en aquests reculls s'observen al llarg del temps, revisen els aspectes teòrics del canvi lingüístic i també expliquen la diversitat de mètodes que ens permeten de donar-ne compte en una llengua determinada i en diferents àmbits de la llengua.

En aquest llibre també s'integra una visió del canvi lingüístic des de la teoria de la complexitat (complèxica), especialment en les intervencions de Vicent Salvador, Àngels Massip, Gemma Bel-Enguix i Emília López-Iñesta i Ernest Querol (UOC) (amb Francisco Grimaldo i Manel Perucho).

Vicent Salvador en la seva contribució «Toc toc —qui és?—. El canvi (reflexions pragmadiscursives)» exposa unes reflexions estimulants sobre el canvi lingüístic, com ara la consideració de l'escriptura com una innovació revolucionària tant des del punt de vista cognitiu —dur a terme l'activitat lingüística que és eminentment oral, amb la mirada— com des del cultural. Revisa les concepcions del canvi i els enfocaments d'estudi en les diferents teories lingüístiques. Relaciona el canvi amb la variació. I considera que el lingüista que vol acostar-se a l'examen del canvi ha d'adoptar una perspectiva àmplia que excedeix el «sistema» de la llengua i que, en bona mesura, integra factors contextuais imprescindibles. En aquest sentit, remarca el marc de la complexitat com a idoni per a l'estudi dels sistemes complexos que es basen en l'adaptació *ecològica* a situacions canviants, de manera que els contextos no solament són fonamentals per entendre el sistema sinó que *esdevenen part del sistema*.

Àngels Massip en «Variació lingüística i canvi: aproximació des de la perspectiva dels sistemes complexos adaptatius» proporciona una introducció teòrica general al tema del canvi lingüístic des de la perspectiva dels sistemes dinàmics i aplica els conceptes desenvolupats pel paradigma del grup ECCO (Evolució, Complexitat i Cognició) com un marc conceptual basat en una ontologia de l'acció.

Irene Castellón, en la seva contribució «Mètodes de processament per a l'anàlisi lingüística de la variació» fa una revisió dels diferents mètodes i eines de la lingüística computacional que s'han desenvolupat en els darrers anys i que poden ser útils per a l'anàlisi lingüística de la variació. Fa èmfasi en l'existència de sistemes d'anàlisi lingüística que ens permeten etiquetar textos de forma automàtica molt ràpidament, amb la qual cosa es pot disposar de moltes dades anotades. També parla d'una altra línia d'investigació, la dels sistemes inductius, que són capaços de construir models de llenguatge només a partir de les dades. En realitat, per al desenvolupament d'aquests sistemes calen corpus; en alguns casos, els corpus poden ser col·leccions de textos sense anotar, el que anomenem text pla, o bé textos anotats on la informació s'ha fet explícita. Es tracta de sistemes que poden guardar informació de forma distribuïda i poden crear la seva pròpia autoorganització de la informació. Aquests sistemes permeten resoldre problemes per a tasques de classificació, de reconeixement de patrons i de reconeixement de comportaments.

Gemma Bel-Enguix en el seu capítol «Cap a una quantificació de l'evolució semàntica del lèxic mitjançant tècniques basades en corpus diacrònics i word2vec» proposa un estudi sincrònic basat en corpus de diferents períodes per fer-ne després una interpretació diacrònica. I això, seguint els corrents actuals d'estudi del llenguatge natural, que es basen en la compilació de grans corpus de text i l'anàlisi mitjançant mètodes estadístics, i que tenen com a idea central que això pot servir per interpretar des d'un altre punt de vista el tema de l'evolució del llenguatge. S'hi explica l'aplicació inicial d'algunes tècniques (anàlisi de freqüències, *word embeddings*, vectors entrenats amb xarxes neurals —com word2vec i altres sistemes de vectors— a un corpus diacrònic català. D'aquesta manera, es vol comprovar la idoneïtat de la metodologia en recursos i llengües diferents de l'anglès. Concretament, en la secció 2 es fa un breu repàs del marc general d'estudi del canvi semàntic dins del processament del llenguatge natural. La secció 3 se centra en l'explicació general de la metodologia emprada en la preparació dels corpus i la generació dels vectors. En la secció 4 s'expliquen els experiments basats en freqüències, mentre que la secció 5 desenvolupa el tema dels experiments basats en vectors. Finalment, en la secció 6 es presenten les conclusions i direccions futures de treball.

També han intervingut en el col·loqui Ernest Querol i Emília López-Iñesta, en representació del grup del qual formen part també Francisco Grimaldo i Manel Perucho, amb «La predicció del canvi lingüístic mitjançant la classificació supervisada», on fan una anàlisi de les modelitzacions proposades en la sociologia primer i en la lingüística (sociolingüística i sociologia del llenguatge) després. Ponderen l'avantatge de fer servir la ciència de la complexitat

perquè es pot connectar la recerca lingüística, la sociolingüística variacionista o la sociologia del llenguatge amb molts altres camps que estudien els sistemes complexos, i adoptar-hi els principis i els mètodes que han estat satisfactoris en altres disciplines. Hi presenten algunes recerques elaborades pel Grup de Recerca Alcàntera: el Pont amb la Sociolingüística (GRA), que s'inscriuen en l'anàlisi dels processos de substitució i de reversió lingüística.

I Maria-Pilar Perea tanca aquest volum amb el seu treball «Canvi lingüístic i anivellament dialectal». L'objectiu del seu capítol, que se centra en l'estudi de la morfologia verbal del català occidental (o nord-occidental), del valencià i del balear, pretén, d'una banda, donar compte dels processos de canvi que han afectat els diferents morfs (i de vegades les arrels) que componen els verbs que paradigmàticament s'han considerat models regulars; i, de l'altra, descriure la naturalesa del canvi i, en la mesura que es pugui, determinar-ne la motivació.

Aquest llibre, doncs, aporta noves maneres de prendre en consideració el canvi lingüístic i integra diferents mètodes i eines per observar-lo i estudiar-lo, així com aplicacions concretes a la llengua catalana, per la qual cosa confiem que constitueixi un instrument per orientar els futurs estudis sobre el canvi lingüístic en l'ampli ventall de llengües que s'estudien a la nostra Facultat.

ÀNGELS MASSIP-BONET
MARIA-PILAR PEREA
Universitat de Barcelona

Barcelona, setembre del 2022

Toc toc —qui és?—. El canvi (reflexions pragmadiscursives)

Vicent Salvador

Universitat Jaume I de Castelló

La primavera ha venido,
nadie sabe cómo ha sido.

ANTONIO MACHADO

Sento les fulles caure
d'octubre, lentament,
roges, i l'heura verda
tornar-se groga al mur
del meu opac silenci.

JOAN VINYOLI

1. PREÀMBUL

En primer lloc he de confessar que el meu paper en aquestencontre sobre el canvi lingüístic és, en la intenció dels organitzadors, el de projectar-hi una mirada de foraster, d'algú que veu les coses des d'un altre àmbit aliè al de la lingüística històrica i les seues metodologies. Per això, com a cautela i advertència, he volgut posar a aquesta intervenció un títol informal, d'anar per casa, tot i que matisat per un subtítol de to més acadèmic a tall de coartada. A més, les citacions poètiques que encapçalan el text són indicis d'un cert propòsit de reflexió d'ampli abast sobre els canvis de tota mena en l'esdevenir de la naturalesa i de la vida humana.

Certament, la recerca que he fet sobre aquesta àrea és ben minsa, i pràcticament es redueix a la diacronia de les oracions concessives, temàtica que és el meu principal lligam amb la sintaxi (Pérez i Salvador, 2014). Tanmateix, el meu itinerari com a investigador en el camp de l'estilística de base pragmàtica és més prolix, més reincident, com es palesa en una altra intervenció feta en aquest mateix fòrum del CLUB fa un grapat d'anys (Salvador, 2013), i s'estén fins a treballs més recents (Salvador, 2020; Payrató i Salvador, 2021).

Dic açò perquè he reivindicat sovint que la variació estilística és un dels mecanismes del canvi en les llengües. A fi de comptes, un dels axiomes d'una

lingüística funcionalista —enfront d'altres perspectives com ara la del generativista— és el seu compromís amb un model basat en l'ús: l'estudi del que els parlants fan amb la llengua, i no pas del que coneixen. No cal dir que aquesta anàlisi dels usos lingüístics no es redueix als mètodes de la lingüística de corpus, però és innegable que aquesta pot ser una eina de treball productiva. En tot cas, amb aquesta mirada que podem qualificar d'externa sembla que la variació funcional escorcolla més els usos socials que el funcionament psicològic individual de la ment, i en aquest sentit es lliga íntimament amb el fenomen del canvi lingüístic. Podria adduir ara certs plantejaments de Paul Hopper quan postula que la lingüística *stricto sensu* podria considerar-se com una mena de microretòrica, és a dir, un estudi de les microestructures del que seria la retòrica en un sentit ampli; tot això, en la perspectiva de les gramàtiques emergents (Hopper, 2007 i 2012). És des d'aquest punt de vista que he gosat acostar-me a la temàtica d'aquesta iniciativa del Club.

2. LA HUMANITAT EN LA SEUA HISTÒRIA

La dialèctica entre permanència i canvi és una de les claus més pregonades de l'existència humana. I aquesta dialèctica és viscuda sovint com un enigma, un misteri que la literatura ha registrat manta volta amb l'ajut de la paraula poètica. Certament, la consciència, com a valor diferencial de l'ésser humà, situa aquest ésser humà en una cruïlla de les coses que perduren, aparentment immutables, i, de l'altre costat, la seva pròpia trajectòria, irreversible, entre el naixement i la mort: entre un no-res i un altre, vaja. Els cicles de la naturalesa, com ara les estacions de l'any indefinidament repetides, contrasten amb la vectorialitat orientada i irrepetible de la vida humana. Això no és una vivència merament personal, sinó que es trasllada a les diferents visions culturals. En efecte, hi ha una frontera cognitiva entre la concepció panteista del món com a etern retorn i la concepció judeocristiana —i marxista— de la història com a itinerari de progrés cap a una meta d'arribada. De fet, sense aquesta segona perspectiva no existiria la història com a disciplina i tot seria mentalitat mítica. Hi ha el canvi, en tot cas, i sovint ens sorprèn, sia perquè modifica la configuració del tauler de joc de manera sobtada —catastròfica, en el sentit matemàtic que va introduir René Thom—, sia perquè s'ha apropat imperceptiblement, de puntetes, com un lladre expert en la intromissió del qual no som conscients fins que descobrim que algú ens ha furtat els mobles i el rellotge de carilló que havíem heretat de l'àvia.

La recerca de les ciències del món físic sol operar sota el pressupòsit que les lleis de la física són inalterables, es consideren congelades en el temps i idènti-

ques en l'espai, o que almenys l'eventual canvi no seria observable a escala humana, ni en una trajectòria individual ni tan sols en un llarg període històric. Tanmateix, cap geòleg o cap naturalista, ni, per descomptat, cap historiador, podria fer aquesta suspensió dels processos temporals, aquesta reducció eidètica que posa entre parèntesis els canvis. Tant els processos d'erosió constant com els terratrèmols planetaris, l'evolució de les espècies i les transformacions històriques, són essencials en aquestes disciplines. La potència intel·lectual de Darwin va situar alguns d'aquests fenòmens en un marc sistemàtic i va marcar l'agenda intel·lectual dels estudiosos. També dels lingüistes, no cal dir-ho. El canvi és, a partir d'aquestes coordenades, un fenomen teoritzable, en el qual no cal suposar cap mena de teleologia, sinó l'aplicació automàtica d'unes lleis, com ara les de la selecció natural, en interacció adaptativa amb el medi.

Així, tal com cada moviment en el tauler d'escacs pot reconfigurar la partida, bé de manera poc rellevant, bé de manera decisiva, els canvis en la història de la humanitat o en l'evolució d'una llengua han de contemplar-se de manera adaptativa, com a cerca d'estratègies en una cadena dinàmica de *problem-solving*. L'estructuralisme ens va ensenyar la productivitat epistemològica d'aquesta concepció dels desequilibris i reequilibris del sistema a través de talls diacrònics. El funcionalisme i l'estudi del discurs han ampliat la dimensió del context en què aquests processos tenen lloc, com veurem suara.

En un text recent, Sebastià Serrano (2016) va ponderar la importància històrica de l'aparició de l'escriptura des del punt de vista de l'antropologia i de la neurologia:

Aquells escribes que van introduir l'escriptura van fer una innovació que a hores d'ara em semblaria, tant des del punt de vista científic com fins i tot des d'altres punts de vista, un tipus d'innovació extraordinàriament agosarada. El que pretenien és que poguéssim realitzar activitat lingüística amb els ulls, que poguéssim realitzar activitat lingüística amb la mirada. Vet aquí que hauríem de dir que l'evolució del llenguatge en cap moment ens hauria fet pensar —després de reflexionar molt sobre aquesta evolució— que era impossible, o gairebé, implicar el sistema visual en la seva producció i en la seva recepció.

En efecte, la gran aportació dels primers escribes a la comunitat humana universal —el gran canvi innovador— va ser la connexió del funcionament cerebral de tipus auditiu, via pròpia del llenguatge natural, amb la via del canal visual, regit per principis neurològics radicalment diferents. La introducció de l'escriptura, en aquest sentit, va determinar una autèntica revolució cognitiva i cultural de la qual encara som deutors.

3. LA POLISÈMIA DEL MOT «CANVI»

Una senzilla observació de certs usos de la paraula «canvi» en català ens pot guiar en la reflexió sobre els seus significats, emparentats però diversos. Compararem ara tres locucions que contenen el mot: «en canvi», «a canvi de» i «canvi a» (o «canvi de»).

- 1) D'aquesta manera, doncs, el connector «en canvi» estableix una relació de contrast entre dos enunciats, de manera equivalent a un connector adversatiu precedit generalment d'una pausa: «X, en canvi Y» («a tu t'agrada el teatre, en canvi jo preferesc el cinema»; «aquest partit és una mica xenòfob; en canvi l'altre no ho és»). En aquest tipus de construccions seria commutable per un «però» o, més freqüentment, per «mentre que». Per això el seu ús implica convencionalment contrast i, si els dos membres connectats no són susceptibles d'establir contrast pragmasemàntic, l'oració és molt difícil d'interpretar o bé palesa ironia per part de l'enunciador (com ara aquesta *boutade*: «certament duc malament el pas dels anys; en canvi tu... també»). Fora d'aquest tipus de construccions, el mot «canvi» no manifesta directament un contrast, però sí que aquest pot ser un dels seus components implícits, com veiem en els apartats següents.
- 2) Un segon ús del mot en unes altres locucions connectives és el d'expressar la idea d'intercanvi entre dos elements diferents però equivalents quant al seu valor dins un marc pragmàtic determinat: «li va donar un feix de bitllets a canvi del seu silenci». El seu contingut semàntic pot ser caracteritzat com un *do ut des*, en contextos com ara l'intercanvi de favors, de parelles o l'estival entre dos joves de països diferents a fi d'aprendre idiomes estrangers, per exemple. Es juga ací entre la diferència (una mena de contrast) i la d'equivalència pragmàtica. L'expressió «barat a» seria un substitut adequat en molts contextos.
- 3) El tercer valor del mot en expressions com ara «canvi a» o «canvi de», que és el més ens interessa ací, conté també una idea de diferència i de substitució d'un element per un altre, però a més ens trobem amb un procés cronològic unidireccional que connecta dos estats de coses successius en el temps i que, prototípicament, és irreversible. Equival al significat de convertir-se, tornar(-se), transformar(-se), mudar, esdevenir, evolucionar... Les construccions d'aquesta mena poden contenir referències explícites als estats cronològics i a la meta de la transformació («ha canviat a millor», «s'ha canviat al barri de Salamanca») o bé poden focalitzar l'àmbit nocional en el qual es produeix el canvi («ha canviat d'idees», «vol canviar de parella»).

O pot també expressar-se d'una manera més sintètica encara, amb un sintagma nominal en què l'adjectiu indica l'àmbit on es produeix la transformació: «canvi climàtic». En el context de la biologia es parla a vegades de «mutació» de gens com a alteració sobtada i permanent que fins i tot pot passar a la descendència d'un ésser viu.

En síntesi: si el primer ús que hem examinat (1) es limita a focalitzar la diferència (contrast) entre dos elements comparables i el segon (2), l'equivalència entre elements diferents, el tercer (3) implica transformació, és a dir, evolució en el temps. Quan parlem de «canvi climàtic» o de «canvi lingüístic» es tracta d'aquest darrer significat (mentre que en el primer d'aquests exemples, les connotacions són negatives, disfòriques, en el segon són neutres o en alguna ocasió positives, tal com s'esdevé majoritàriament en el «canvi polític» que molts partits volen proposar en els seus programes electorals). Tot això amb el benentès que qualsevol canvi d'un element implica una modificació en la corresponent configuració de conjunt i, per tant, un trànsit entre dos estats de coses successius que són diferents entre si.

4. ELS TIPUS, LES CAUSES I ELS MECANISMES DEL CANVI LINGÜÍSTIC

Els manuals de lingüística històrica precisen bé la tipologia d'aquest canvi, és a dir, quins components o plans de la llengua es veuen afectats en cada cas per aquestes transformacions: el component fònic, el morfològic, el sintàctic, el lexicosemàntic... Els manuals també presenten les causes que provoquen aquestes evolucions: causes endògenes, com ara l'analogia, la reanàlisi (o neoanàlisi), la metaforització, la metonímia o l'eufemisme; i exògenes, com són els préstecs, les interferències interlingüístiques, els neologismes, que designen noves realitats, o bé els factors sociolingüístics, que privilegien unes variants sobre altres per motius de prestigi. (Les metàfores i els eufemismes tenen generalment una mena de data de caducitat, perquè es desgasten com a portadors de significats indirectes i les expressions corresponents es convencionalitzen i es fan més transparents en un procés de lexicalització.)

D'altra banda, els mecanismes del procés de canvi acostumen a operar unidireccionalment en el sentit d'incrementar la subjectivitat, i també anant d'allò més tangible a allò més abstracte: els continguts espacials, així, poden evolucionar cap a un valor temporal (com, per exemple, en el cas de l'espanyol «luego») i, d'aquest valor temporal, cap a un sentit més abstracte, com ara el causa/con-

seqüència ('llavors' o 'aleshores', per exemple, en aplicació del principi filosòfic que predica *post hoc ergo propter hoc*). Les accions humanes donen pas a processos asèptics, com quan la percepció de culpa esdevé simplement causa. Es considera, així mateix, el fet que el canvi sol ser molt lent, sovint imperceptible, en el transcurs d'unes poques generacions, o, a vegades, abrupte en el nivell de la semàntica lèxica, quan el temps històric s'accelera (per exemple, en el camp de les tecnologies modernes o en la revolució dels costums). Els casos de reanàlisi, per exemple, són difícilment percebuts pels estudiosos de la lingüística històrica quant a la cronologia del procés.

Tot això és ben conegut i no manquen en els estudis de les darreres dècades les reflexions sobre certs factors pragmadiscursius, com veurem més endavant. Durant aquests anys s'ha insistit en la interfície pragmàtica/gramàtica, en el procés de convencionalització com a estructures gramaticals de les estratègies més reeixides en la pragmàtica del discurs, i en conceptes com ara la (inter)subjectivització (Traugott 2010 i 2014), el coneixement compartit pels interlocutors o la «inferència invitada» (Traugott i Dasher, 2002; Antolí Martínez, 2019).

Sembla obvi que el procés de canvi va lligat a la variació: la diatràtica, la diastràtica o la funcional, per mitjà de mecanismes complexos que poden combinar aquestes diferents modalitats de la variació sincrònica d'una llengua. El canvi diacrònic parteix de l'heterogeneïtat dels usos que s'observen en un estat sincrònic i progressa d'una manera gradual que sovint resulta poc perceptible al llarg d'un contínuum temporal (Bermúdez-Otero i Trousdale, 2012). Les pautes de prestigi hi operen sociològicament. Com se sap, William Labov va explorar alguns d'aquests mecanismes i va mostrar com actuen aquestes pautes en l'anglès americà.

5. UN ESTUDI DE CAS: LA VARIACIÓ DEL PERFET CATALÀ

Un cas del català que ha estat estudiat per diversos lingüistes —Colón (1978 i 1987), Salvador (1995a), Pérez Saldanya (1996), Pérez Saldanya i Hualde (2003)— és el de la variable *perfet simple/perfet perifràstic* en català, que ens permet observar un canvi lingüístic en marxa durant molt de temps fins avui mateix. Germà Colón va establir amb claredat el marc d'estudi i la manera com el canvi, sobre la base d'una variació geogràfica, s'inicia en un determinat gènere medieval i en certs cotextos: passatges on la narració empra el present històric per donar color al relat i recorre al verb «anar», que comença a dessemantitzar-se. Es produeix, tal com ha estudiat Pérez Saldanya, una combina-

ció de gramaticalització i reanàlisi. Més tard, aquesta construcció del VADO + infinitiu va alliberant-se d'aquestes restriccions co(n)textuals i del marc genèric o tradició discursiva —a narració èpica— i es generalitza, molt més del que ho fa en altres llengües romàniques on no va arribar a qualla com a estructura gramaticalitzada. Paral·lelament, allò que les *Regles d'esquivar vocables o mots grossers o pagesívols* consideraven vulgar i propi dels parlants de les muntanyes (la variant «vaig parlar» enfront de l'alternativa culta «parlí») s'imposa per un canvi de les pautes de prestigi, mentre que el perfet simple només sobreviu —i ho fa reculant davant l'altra alternativa, sobretot en la forma de primera persona del singular— en dialectes geogràfics com ara el valencià central i, això sí, en molts textos escrits de caràcter culte, on, a més, els escriptors poden aprofitar l'estratègia sinonímica corresponent a aquesta alternança possible a fi d'evitar repeticions enutjoses.

Per altra banda, el valor narratiu de la forma del perfet simple fa que el seu ús siga percebut com a menys anòmal en la forma de tercera persona, mentre que la de la primera («vinguí», «fiu», sovint amb formes irregulars) és sentida per la majoria de parlants del domini lingüístic català com a anòmala i es converteix en una marca estereotípica que pot tenir un valor negatiu per la seua raresa (de fet, pot suscitar fàcilment la caricatura humorística), però també constitueix, alhora, una senya d'identitat comarcal assumida pels parlants en certs casos: concretament, en el valencià central, on té encara una forta vitalitat, si més no en la parla col·loquial.

D'alguna manera, doncs, la variable pot activar els tres tipus assenyalats per Labov: *indicador* (procedència dialectal), *marcador* (del canvi de registre o de gènere) i *estereotip* (variant conscient i notòria amb una funció de reforçament de la identitat grupal). Es pot observar el funcionament d'aquesta variable en l'estilística dels textos literaris, en concret en el d'alguns escriptors valencians. En efecte, d'una banda, el manteniment de les formes simples en el parlar col·loquial d'aquestes àrees coincideix plenament amb la tradició literària clàssica i és, així mateix, l'opció preferida en la producció valenciana de l'època moderna, sovint amb caires popularistes, des de Pere Jacint Morlà (segle XVII) a la *Rondalla de rondalles* (segle XVIII) o als col·loquiers. Ara bé, quan a la segona meitat del segle XIX hi ha una clara preferència per la forma perifràstica en la major part del domini lingüístic, el sainetisme d'Eduard Escalante continua emprant de manera predominant la forma simple (tot i que ho fa més matisadament en la primera persona, on sovint alterna amb la perifràstica), en contrast amb el cas coetani de Pitarra (Salvador, 2001: 63-64).

En qualsevol cas, fins i tot en el «català que ara es parla», tal com el va voler incorporar Pitarra a la seua escriptura teatral, no deixen de trobar-se exemples

d'aquest ús de les formes simples. Però podem analitzar amb profit com operen, en aquest camp, els autors valencians contemporanis, concretament pel que fa al cas de Vicent Andrés Estellés. El poeta de Burjassot va aconseguir, en la seua llengua poètica, un equilibri encomiable entre el català estàndard i un regust estilístic de localisme valencià. En aquest sentit, Estellés empra usualment la morfologia valenciana compatible amb la normativa, per exemple en les formes verbals («jo plore», «que ella vinga»), en els possessius («meua», «teua») i en alguns particularismes lèxics. A més, potencia la fraseologia i la paremiologia, que prenen sovint un cert color local («tocar mare», «qui no suma se n'ix», «caldera vella, bony o forat»), tot i que, com defensa Maria Conca (2009), la major part de les unitats fraseològiques que empra no s'allunyen gaire del català general i, en tot cas, formen part de l'arxiu fraseològic de la llengua. Ara bé, en el cas de la variable que estem analitzant, el poeta fa predominar la variant perifràstica, sobretot pel que fa a la forma de la primera persona. El resultat és un equilibri admirable en un estil apropiat a la parla popular en molts aspectes però alhora ennoblit per la fidelitat al català estàndard i per altres recursos, com ara les intertextualitats amb la poesia clàssica, molt particularment la d'Ausiàs March.

En canvi, un narrador valencià com és Josep Lozano, en la seua novel·la *Ribera* —un homenatge a la seua comarca, amb una habilitíssima utilització de nombrosos fraseologismes, sovint extrets de la parla local—, assaja l'ús molt freqüent del perfet simple, fins i tot en la forma de la primera persona. Òbviament, es tracta així d'exhibir orgullosament un tret dialectal com a emblema d'identitat dialectal.

Tot plegat, el que observem en aquests escorcolls literaris és que la variable que he comentat juga, en la literatura valenciana contemporània, amb tres pautes de prestigi tan diferents com legítimes: a) la tendència uniformitzadora de l'estàndard a favor de la forma perifràstica; b) la invocació, en alguns casos, del prestigi de la llengua clàssica (a favor de la forma simple); i c) el valor identitari, per a una collectivitat comarcal, de la variant avui considerada dialectal (perfet simple). La variació lingüística es palesa així no sols com l'origen del canvi lingüístic, sinó com una persistència actual que no deixa d'oferir possibilitats expressives a l'estilística literària.

Una reflexió final sobre aquest tema ens duria a considerar la diferència entre el procés de canvi lingüístic en general i el canvi en una esfera socio-semiòtica molt diferent com és el funcionament de les comunitats científiques. En aquests darrers àmbits, l'evolució tendeix a arribar a nous consensos per mitjà del que s'anomenen «revolucions científiques», en termes de Kuhn. Per a aquest autor, els paradigmes científics són més resistents a les anomalies obser-

vades del que preveia el plantejament més lineal de tipus popperià, però tan bon punt les anomalies es multipliquen i emergeix un nou paradigma en la ciència de què es tracte (tot i que això no és aplicable tan directament a les ciències humanes, on pot conviure més d'un paradigma simultàniament), es produeix la «revolució», que, en un moviment dialèctic, dona a llum un nou paradigma i la comunitat científica tendeix a establir un ferm i definit consens al voltant d'un nou paradigma. Pel que fa al canvi lingüístic, hi ha l'emergència d'innovacions, com hem vist, però la llengua conserva moltes vegades diverses variants de la variable corresponent en un mateix tall sincrònic.

6. LA INTERFERÈNCIA PRAGMÀTICA COM A RUTA CAP AL CANVI SINTÀCTIC

És clar que molts canvis no responen a mecanismes interns de l'evolució d'una llengua determinada presa com a entitat aïllada, sinó a influències (préstecs, calcs, interferències) d'una altra. Hi ha bones raons per pensar que en general les estructures sintàctiques són més resistents al canvi que, per exemple, les unitats lèxiques aïllades, que s'innoven amb més rapidesa per préstec o per neologia. És curiós observar que aquests canvis sintàctics es poden produir originàriament per la transformació dels contextos d'ús quan aquests s'adopten per l'influx d'una llengua diferent. Aquest és el cas de la interferència pragmàtica, és a dir, quan una estructura perdura formalment inalterada durant un temps però els seus usos discursius —ço és, el seu comportament contextual— s'alteren per influència d'una altra llengua amb la qual té contacte continuat, i sobretot si aquesta és una llengua de més prestigi, com en el cas de l'anglès respecte a l'espanyol o al català (i fins i tot el de l'anglès respecte al català a través de l'espanyol). En efecte, les alteracions graduals dels usos discursius produeixen un trasbals de les funcions sintàctiques i poden arribar a provocar canvis formals o reanàlisis de les estructures anteriors. Caldria afegir que, òbviament, en la lingüística anglòfona no hi ha tant d'interès en l'estudi d'aquests fenòmens com quan el canvi opera des de l'àmbit de l'anglès, més prestigiat que el que rep la interferència, cap al receptor de la influència.

En un ordre de coses més general que el del canvi lingüístic, podem visualitzar i conceptualitzar aquest procés per analogia amb un mecanisme de canvi anatòmic en les persones. Així doncs, una persona que té dolor físic en una determinada part del cos com ara l'esquena o un peu, força la seua conducta postural (camina o s'asseu freqüentment d'una altra manera, per exemple) i aquest hàbit de contrarestar el dolor modifica la funció d'una part contigua del

cos i finalment pot produir una lesió o una deformació anatòmica. Una persona pot, així, encorbar-se d'espatla a força d'anar modificant la seua postura durant molt de temps per adaptar-se a una afecció dolorosa i reduir el seu patiment. El canvi és així poc perceptible i passa per una modificació progressiva de les funcions realitzades habitualment per alguns músculs, cosa que pot derivar en forts canvis anatòmics (la gepa, per exemple, o alguna mena de coixesa). En tot cas, si analitzem l'anatomia d'un cos en dos períodes diferents, podrem copsar perfectament el canvi produït, i observant el procés com si ho veiéssim a càmera lenta, comprendrem millor el mecanisme que ha produït la modificació deformadora. Dit sintèticament: les estratègies comportamentals —en el nostre cas, les de la comunicació discursiva— quan esdevenen un *habitus* produeixen modificacions en les estructures.

Tornant a la lingüística, recordaré que la sociolingüista xilena Carmen Silva-Corvalán (1993) va analitzar alguns casos d'aquesta mena que s'expliquen per interferència pragmàtica de l'anglès sobre l'espanyol, en una línia que vaig continuar personalment en un treball posterior (Salvador, 1995b). Des d'aquest punt de vista, la interferència no consisteix en la introducció d'estructures alienes a la llengua receptora, sinó en una ampliació dels usos contextuals de construccions ja conegudes que experimenten una expansió explicable des d'un plantejament cognitiu, en un procés que pot desembocar en un canvi estructural. S'explica així la tendència de l'espanyol actual a ampliar l'ús dels possessius en casos en què la llengua diana no recorreria a aquesta estructura, sinó, per exemple, a un datiu possessiu o a una mera el·lipsi: «lávese las manos» esdevé, així, «lave *sus* manos», i «guarde el recibo de compra» es converteix en l'expressió, semànticament redundant, «guarde *su* recibo de compra». De la mateixa manera, un enunciat com ara «esta mañana me he encontrado con alguien muy querido para mí», amb aquest «alguien» inespecífic (en comptes d'«un amigo», «una mujer» o «un pariente»), no deixa de suggerir, en espanyol, una certa voluntat d'intriga o, almenys, de reserva d'informació, mentre que l'estructura anglesa de la qual es calca («somebody») és més acceptable en aquesta llengua. Un camí semblant pot observar-se en la interferència de l'espanyol sobre el català en construccions com ara l'acusatiu preposicional o el futur de probabilitat.

7. ELS PROCESSOS DE GRAMATICALITZACIÓ

Les recerques sobre els processos de gramaticalització per part d'autors com ara Givón, Hopper o Traugott, entre d'altres, han projectat una nova llum sobre certs fenòmens de la lingüística històrica. En la base d'aquests processos hi ha

la coneguda referència de Givón al fet que el discurs d'ahir és la sintaxi d'avui, alhora que la sintaxi d'avui és la morfologia de demà. Podem dir que les estratègies discursives, quan són reeixides comunicativament, tendeixen a convencionalitzar-se i poden acabar donant lloc a noves estructures fixades en la llengua, és a dir, poden arribar a *gramaticalitzar-se*, com passa amb molts usos metafòrics de mots o fraseologismes que menen a una lexicalització que fixa noves unitats lèxiques o fa sorgir noves accepcions d'aquestes unitats. Podem considerar que els processos de gramaticalització protagonitzen un trànsit de la retòrica, en un sentit ampli, a la gramàtica (Salvador, 2020).

La història de les construccions concessives, i concretament la dels connectors que les marquen, és llarga i complexa en diverses llengües. En espanyol aquesta història constitueix un extens capítol de la sintaxi històrica on «aunque» va imposar-se com a conjunció prototípica —tal com fa l'eucaliptus depredador amb les altres espècies arbòries que troba al seu entorn pròxim—, tot i que sense anular les altres opcions existents encara en la llengua actual (Pérez i Salvador, 2014). En català, l'evolució d'aquests connectors també ha estat complicada i la prototipicitat de l'«encara que» no ha eliminat tampoc la subsistència de moltes altres opcions ni les diferents construccions amb valor concessiu que, sense necessitat d'un connector canònic, se senten avui com convencionalitzades en un grau o altre i —podríem dir— *semigramaticalitzades*. En el fons es tracta d'un pas intermedi entre, d'una banda, construccions que generen inferències contextuais i, de l'altra, l'emergència d'una mena de connectors que es van convencionalitzant i que són susceptibles de gramaticalitzar-se, tal com va succeir, en èpoques anteriors, amb els que avui són connectors concessius de ple dret (Salvador, 2010). Vet ací uns pocs exemples d'aquest tipus de construccions que palesen fàcilment un valor concessiu:

- (1) *Ni que* em matin aniré a l'espectacle.
- (2) No m'ha agradat la pel·lícula, *i això que* tenia ganes d'anar al cinema.
- (3) *Sent com* és un advocat famós, avui no li ha lluit el pèl.
- (4) *Tan* llest com és i no ha pogut resoldre l'endevinalla.
- (5) *Ja pots* predicar a aquesta gent, *que* no t'han de fer cas.

Si bé es mira, les estructures concessives conserven en bona mesura les petges dels mecanismes interactius (conversacionals) on es van originar com a estratègies, en una mena d'interfície entre pragmasintaxi i morfosintaxi en la

qual tenen lloc els fenòmens de polifonia enunciativa (Couper-Kuhlen i Thomson, 2000). Al capdavant, en una visió actual, tot discurs prové de la interacció, i aquesta interacció és pautaada per marcadors que senyalitzen el flux de la conversa, com ara quan «a tot açò», «per cert» o «a propòsit» indiquen que un dels interlocutors vol canviar de tema o vol reconduir la interacció mateixa. Podríem dir, en casos com aquests, que hi ha una mena de recurs «pragma-sintàctic» que relaciona accions discursives entre si.

En les darreres dècades s'ha produït un contacte fructífer entre la teoria de la gramaticalització i la *gramàtica de construccions* que va originar-se principalment al voltant de lingüistes de la Universitat de Berkeley com ara Charles Fillmore, Paul Kay, George Lakoff o, posteriorment, Adele Goldberg, i que ha tingut un desenvolupament continu fins avui mateix. El fet és que, des de primeries dels anys 2000, alguns investigadors de la *gramàtica de construccions* van interessar-se per la dimensió diacrònica de les construccions gramaticals. Els estudis sobre la gramaticalització, que tenien el seu origen en els clàssics treballs d'Antoine Meillet i que van establir-se com a paradigma modern a partir de 1993 gràcies a l'aparició de dos manuals ben coneguts, han rebut la influència de l'altra teorització i han mostrat una notable productivitat epistemològica i metodològica.

Es pot comprovar que una part important de la bibliografia recent es focalitza precisament en els resultats d'aquesta opció que correspon a la *gramàtica de construccions diacrònica* en relació amb la teoria de la gramaticalització (Barodal *et al.*, eds., 2015). Així ho fa, per exemple, el volum que duu un títol ben il·lustratiu: *Grammaticalization meets construction grammar*. La gramàtica de construccions diacrònica elabora una història lingüística constructivista que aborda l'estudi de la diacronia amb una perspectiva pretesament més àmplia que la de la teoria de la gramaticalització, amb un enfocament cognitivista i que inclou el préstec de construccions gramaticals entre llengües (Noël, 2007). Fins i tot en alguna ocasió s'ha proposat substituir el nom *gramaticalització* per *construccionalització*, però no hi ha consens sobre la identificació d'ambdós processos. Probablement els estudis sobre la gramaticalització han ampliat excessivament el seu objecte d'estudi, cosa que ha produït una certa inflació en el concepte, que corre el risc de ser massa inclusiu (Mengden i Simon, 2014). Siga com siga, hi ha un extens consens sobre l'interès d'aquesta interconnexió, malgrat que pot trobar-s'hi algunes incompatibilitats en àrees menors (Coussé *et al.*, eds., 2018).

No entraré ara en aquesta discussió teòrica, però el més interessant és que aquest contacte amb la gramàtica de construccions ha fet una aportació innegable als estudis sobre la gramaticalització: la tradició de Meillet (i de Kurylo-

wicz també), en el sentit de prendre com a punt de partida dels processos gramaticalitzadors una unitat lèxica o morfològica individual (el «morpheme-based approach»), dona pas a una concepció més sintagmàtica o contextual d'aquest punt de partença del procés. Com expliquen Mengden i Simon (2014), en lloc de mots aïllats, ara es considera que la gramaticalització requereix la col·locació sintagmàtica amb altres elements, és a dir, el mot o morfema ha de ser inserit sempre en un marc contextual. Dit d'una altra manera: el marc sintagmàtic on s'insereix la unitat lingüística candidata a la gramaticalització no és un mer acompanyament prescindible, sinó un factor absolutament determinant per a l'inici del procés. Es tracta del que podem anomenar «dimensió sintagmàtica» de la gramaticalització: l'associació originària d'un significat amb una determinada configuració (morfo)sintàctica (Noël, 2013).

Diversos estudis han adoptat aquest enfocament per explicar la gramaticalització del verb anglès *have* com a auxiliar per als temps compostos a partir d'uns entorns sintagmàtics on s'acompanya d'un participi; o el valor de futur que assolirà la perífrasi *going to* a partir del significat de moviment, però en casos en què a la preposició *to* segueix un verb i no pas un nom de lloc; o la gramaticalització del verb espanyol *poner* com a auxiliar incoatiu («ponerse a llorar») a partir de configuracions sintagmàtiques concretes. Molts dels exemples estudiats poden trobar paral·lelismes en l'àmbit de la lingüística històrica del català, com ara pel que fa a la construcció *VADO* + infinitiu, tot i que en aquest cas, com hem vist, la construcció acaba assolint un significat de passat en comptes del de futur, però això és explicable precisament per l'entorn inicial, on el verb tenia un valor de present històric i, per tant, s'emmarcava en un context de narració de fets pretèrits. Podríem exemplificar també aquesta estratègia analítica amb les unitats lèxiques inicials de la negació discontinua del català (o del francès), on la gramaticalització amb sentit negatiu de mots com ara *gaire*, *gens*, *res* o *cap*, entre altres, no és explicable si no es considera la configuració sintagmàtica on apareixien inicialment.

8. CODA

Una observació final. Tota aquesta problemàtica que se suscita, com hem vist, en tractar els envitricolls del canvi lingüístic —un fenomen tan suggestiu com complicat, igual que tots els processos de canvi—, troba un marc interdisciplinari adient en la teoria de la complexitat, que també té un lloc en aquesta jornada. Per exemple, pel que fa a la centralitat que el concepte de context assoleix en aquesta teoria. Al capdavant, prescindir del context en lingüística

seria comparable a un tractament de la pandèmia actual a partir només de l'estudi del coronavirus que la causa, tot prescindint de factors de tipus contextual com són les vies socials de contagi, les conseqüències econòmiques i institucionals, la representació de la pandèmia en l'imaginari col·lectiu o la posició dels negacionistes pel que fa a les vacunes... El lingüista que vol acostar-se a l'examen del canvi ha d'adoptar una perspectiva àmplia que excedeix el «sistema» de la llengua i que, en gran manera, integra factors contextuais imprescindibles. La teoria de la complexitat ens proveeix d'un marc molt aprofitable en aquesta tasca: en efecte, el context és cabdal per a uns sistemes complexos que es basen en l'adaptació *ecològica* a situacions canviants, de manera que aquests contextos *esdevenen part del sistema*. Però de tot això tractarà una veu més autoritzada que la meua en un altre capítol.

Variació lingüística i canvi: aproximació des de la perspectiva dels sistemes complexos adaptatius

Àngels Massip-Bonet
Universitat de Barcelona

1. SISTEMA LINGÜÍSTIC, SISTEMA DINÀMIC

Un sistema lingüístic té un comportament clarament dinàmic: té sentit que l'activitat (llenguatge) d'un sistema complex adaptatiu (en anglès, *complex adaptive system*, CAS) com el cervell humà reflecteixi les propietats d'aquest tipus de sistema. De la mateixa manera que el cervell s'organitza en xarxes neuronals, la llengua també s'organitza en xarxes. La forma en què fem ciència també hauria apropar-se més a la forma en què funcionen els nostres cervells: de manera transdisciplinària (per dir-ho de manera metafòrica) i organitzada en xarxes.

La llengua ofereix les mateixes característiques que tipifiquen els cossos complexos del món real: és obert, sense límits i està ple de parts inestables; té infinita cardinalitat pel que fa als seus components, les seves relacions i les seves configuracions; és dinàmic, metaestable (la metaestabilitat és un estat temporalment estable: l'equilibri es trenca quan hi ha pertorbacions en un sistema, en el punt el sistema canvia a un altre equilibri metaestable) i depèn de la trajectòria; no és lineal, és sensible a les condicions inicials, és exponencialment amplificable i en regions caòtiques; i és emergent, no additiu, no modularitzable i organitzacionalment complex (Andrason, 2014).

Quan es parla de llengua, no ens referim només al sistema lingüístic que es tracta com un objecte d'estudi *per se*, sinó també a l'intercanvi comunicatiu, a la llengua com a modeladora de la cognició, a la llengua com a expressió de la nostra naturalesa (Massip, 2007a: 236). I aquesta concepció és absolutament necessària si volem abastar al més àmplia possible el que implica el canvi lingüístic.

En aquest document vull oferir algunes observacions sobre el canvi lingüístic a través de la perspectiva del que succeeix en els sistemes dinàmics. Per desenvolupar aquestes idees sobre el canvi he revisat la filosofia que sustenta el treball transdisciplinari del grup de recerca Evolució, Complexitat i Cognició (ecco.vub.ac.be). L'objectiu de l'ECCO és construir una cosmovisió coherent

per contrarestar la fragmentació en la ciència i en la societat actual. El marc conceptual adoptat per l'ECCO es basa en una ontologia de l'acció, en la qual els constituents fonamentals de la realitat es consideren les accions i els agents que les produeixen.

Els éssers humans canvien la llengua quan adapten els seus recursos lingüístics a nous contextos. Això significa que un sistema lingüístic i els seus usos són mútuament constitutius.

Aquesta perspectiva dels sistemes dinàmics ens proporciona una manera diferent d'enfocar els problemes lingüístics. Reprenc els conceptes de Heylighen i tracto de donar idees d'aplicació en els camps de la lingüística que em són més familiars o en els quals he investigat poc o molt. És clar que cada lingüista, segons la seva àrea d'estudi, podrà treure'n les seves pròpies conclusions.

La naturalesa dinàmica dels CAS es pot concretar en quatre paràmetres que tots posseeixen. La teoria ha desenvolupat diferents marcs i mètodes per donar compte d'aquests paràmetres. I encara que alguns d'aquests mètodes han sorgit en diferents àrees de coneixement per afrontar reptes específics de cada àrea, tots poden ser transportats i aplicats a qualsevol CAS.

- Són no lineals, de manera que els canvis d'estat del sistema no són directament proporcionals al temps, a l'estat actual del sistema o als canvis que es produeixen en l'entorn. La teoria del caos va ser una de les primeres a repartir el paradigma newtonià (que treballava amb sistemes lineals i sistemes en equilibri).
- S'autoorganitzen i la seva organització global emergeix de les interaccions entre les parts. El model d'investigació *agent-based* ha fet aportacions fonamentals per entendre aquest paràmetre dels sistemes complexos.
- S'adapten a una dinàmica d'acció-reacció tant quan interaccionen els diferents agents (que es mouen entre cooperació i competició) com quan els agents interaccionen amb l'entorn. I evolucionen. L'ecologia ha fet aportacions fonamentals al coneixement d'aquest paràmetre dels sistemes complexos.
- Gairebé tots es poden considerar organitzats en xarxes, on la connectivitat és essencial per entendre'n el funcionament. Aquest paràmetre es pot investigar cada vegada millor gràcies al fet que s'han desenvolupat molt les maneres d'obtenir grans conjunts de dades. La visualització de l'organització de les dades també ajuda a entendre aquest paràmetre dels sistemes complexos.

2. CANVI LINGÜÍSTIC

El canvi és una part essencial del funcionament dels sistemes complexos. El canvi lingüístic es produeix com a resultat de l'aparició constant de fenòmens variacionals resultants de les interaccions entre els individus i l'entorn. Quan una d'aquestes produccions es generalitza, diem que el canvi s'ha consolidat. Sovint, però, les formes preexistents o alternatives a la forma consolidada es mantenen en algunes varietats de la llengua.

Aquí fem un paral·lel amb la genètica, en què els estímuls de canvi poden ser *intrínsecs* (procedents dels gens) o bé *extrínsecs* (produïts per les interaccions amb les cèl·lules veïnes). Si observem el canvi en les paraules de qualsevol llengua, també hauríem de distingir els estímuls *interns* (que ja estan integrats en l'ètim de cada paraula i en la seva evolució segons les lleis de cada llengua al llarg de la història) i els estímuls *externs* (tots els contextos en què qualsevol paraula ha estat usada en el temps i el contacte amb altres llengües i/o varietats en un territori determinat).

En aquest capítol no faré una anàlisi específica dels factors del canvi lingüístic en una llengua específica. Només proporcionaré un acostament als elements involucrats en el canvi en les llengües, fent ús d'una adaptació conceptual del paradigma del grup ECCO al canvi lingüístic.

3. TEORIA DE SISTEMES I COMPLEXITAT

La teoria general de sistemes, vista en part com a precursora de la teoria de la complexitat, tracta els sistemes com a estructures estàtiques i ben definides que es van donar objectivament. No obstant això, aquestes suposicions no funcionen per a sistemes adaptatius complexos, com societats, ments, mercats o llengües. En aquests sistemes, les estructures tendeixen a ser difuses, variables i, en un grau important, subjectives (Heylighen, Cilliers i Gershenson, 2007). En aquest context, hem de parlar d'autoorganització i evolució. L'autoorganització és el procés espontani pel qual els sistemes emergeixen i evolucionen, de manera que es tornen cada vegada més complexos, més adaptables i més sinèrgics (Heylighen 2008).

L'aplicació més reeixida dels sistemes generals a les ciències socials bàsiques és la d'Alfred Kuhn (1975), com veurem en el subapartat 4.2.

L'autoorganització (vegeu la secció 4.6, sobre organització) és l'adaptació mútua i la coevolució dels components inicialment autònoms del sistema, és a dir, els agents (*agent* definit com a productor persistent d'accions). L'autoorga-

nització i l'evolució es veuen com dos aspectes del mateix procés d'adaptació espontània (vist de manera integral). A través de les interaccions, els agents desenvolupen una xarxa de relacions cada vegada més sinèrgiques que coordinen les seves activitats. El sistema evoluciona contínuament (l'evolució es basa en la variació i la selecció internes de sistema) i, per tant, es torna més complex, més adaptable i més sinèrgic (vegeu el subapartat 4.2 a continuació).

La complexitat emergeix en els sistemes a través de canvis incrementals fins i tot quan els canvis no sorgeixen d'un moviment deliberat cap a una meta.

El canvi es produeix de manera contínua a través de l'acumulació quantitativa que condueix a revolucions qualitatives periòdiques. El canvi és continu i indefectible (vegeu el subapartat 5).

4. ONTOLOGIA DE L'ACCIÓ: SISTEMES DINÀMICS

En el seu glossari, Heylighen defineix els conceptes del que s'ha anomenat «ontologia de l'acció» i els agrupa en sis seccions: 1) agents i evolució, 2) cibernètica, 3) cognició, 4) interaccions, 5) sistema i 6) organització. Comentaré alguns dels termes definits per Heylighen, els formularé i els posaré en relació amb el canvi lingüístic.

4.1. Agents i evolució

- *Acció*: canvi en la situació, d'un estat inicial (causa) a un estat posterior (efecte). Una acció és intrínsecament relacional, connecta dues situacions i no pot existir independentment dels fenòmens que vincula entre si.
Qualsevol actuació lingüística pot considerar-se una acció.
- *Estat*: situació del món en un moment determinat, que es defineix pel conjunt de totes les accions que es podrien fer en aquest moment (Turchin, 1993).
- *Agent*: productor d'accions autònom i persistent. Generalment les accions estan adreçades a aconseguir un resultat més que un altre. Mitjançant el llenguatge els agents actuen per aconseguir un objectiu, però normalment no actuen en la llengua en si. Només esporàdicament, si algun aspecte d'una llengua té connotacions extralingüístiques o d'identitat relacionades amb el prestigi (o la manca de prestigi), els agents poden actuar per fer usar o deixar d'usar algunes formes més que d'altres.

Qui són els agents en les llengües preses com a CAS? O bé els parlants (Beckner *et al.*, 2009), o bé, alternativament, els elements del sistema i els parlants (segons Lee *et al.*, 2009).

- *Esdeveniment*: una distinció interessant per a la lingüística, al nostre parer, és la diferència entre *acció* i *esdeveniment*. Una acció és voluntària, però un esdeveniment és una acció no produïda per un agent.

Aquesta distinció és molt útil quan parlem de canvi lingüístic, de variació social i, més específicament, de política lingüística. L'*esdeveniment* correspon a fets externs que poden influir en la llengua, i l'*acció*, a fets específics, com ara prohibicions, lleis contra l'ús d'una llengua, censura, etc.

L'*estat* d'una llengua ve donat per la història d'aquesta llengua, pel conjunt d'accions lingüístiques que els agents han realitzat al llarg de la història.

- *Meta*: un dels estats finals o *atractors* als quals és probable que condueixin les accions d'un agent. Les metes en aquest sentit són implícites la majoria de les vegades: l'agent no té coneixement de quin és l'estat final. En política lingüística és, per descomptat, explícit.

Un atractor també es defineix com una restricció evolucionada espontàniament.

- *Variació*: canvi en curs en l'estat del món causat per les accions.

La variació és sempre fins a cert punt cega, ja que cap agent pot preveure totes les conseqüències de les seves accions.

De vegades hi ha innovacions deliberades, però aquestes han de distingir-se de l'operació inconscient que es produeix, per exemple, en l'aprenentatge bàsic i en les tendències del processament mental que han donat forma a certes regularitats fonològiques, sintàctiques i semàntiques.

- *Selecció natural*: retenció selectiva d'un estat determinat perquè no es produeixin més accions que canviïn aquest estat. Correspon al que s'ha anomenat un *atractor de la dinàmica* o *estat amb alta aptitud física*.

I és en aquest sentit que podem parlar de selecció natural en les llengües. Sovint la pèrdua d'una espècie en el món natural (per selecció natural) s'ha comparat amb la pèrdua d'una llengua. No obstant això, la majoria dels casos de mort de llengües no es deuen a la conjunció de fenòmens naturals (o fins i tot a esdeveniments), sinó a esdeveniments i fins i tot accions de repressió humana d'un grup sobre un altre, a accions humanes de

genocidi i lingüicidi, a accions humanes de voler subordinar (i fins i tot esclavitzar) altres grups humans.

- *Evolució*: canvi dirigit a llarg termini en l'estat del món cap a una aptitud general més gran, que resulta de la interacció de la variació d'estats (en part aleatòria, en part adreçada a un objectiu) i la selecció eventual d'estats amb una aptitud superior.

L'evolució es pot veure com una recerca d'aptitud basada en el procediment prova i error, on la variació produeix les proves i la selecció elimina els errors (Heylighen, 1999a i 2007).

L'evolució en biologia (tant si l'objectiu de la selecció és el gen com si és l'individu en el seu conjunt) s'ha de considerar com la conjunció de dos processos paral·lels: el canvi adaptatiu i l'origen de la diversitat (Rovelli, 2016: 219). No totes les característiques d'un sistema evolucionen a la mateixa velocitat o avancen al mateix temps. En general, un sistema evoluciona cap a una major complexitat i els canvis són irreversibles.

L'origen de la complexitat adaptativa en la llengua és anàleg a l'origen de la complexitat adaptativa en la biologia (Christiansen i Chater, 2008: 505).

Per tant, la complexitat adaptativa de les llengües, molt similar a la de la biologia, sorgeix de la variació lingüística «filtrada» a l'atzar a través de pressions de selecció relacionades amb l'aprenentatge i el processament (Massip, 2013: 47).

Però l'atzar n'és només una de les causes. Al meu entendre, es pot parlar, com en la realitat biològica, de quatre factors: 1) herència; 2) evolució; 3) medi; 4) atzar. I els gens, en l'evolució biològica, són només els prismes a través dels quals es filtra i es refracta l'atzar (Mukherjee, 2017: 257). De la mateixa manera, en les llengües els estímuls intrínsecs (vegeu el subapartat 2) són el prisma a través del qual l'atzar o altres causes es filtren i es refracten. Aquestes altres causes poden ser, entre altres, la facilitat d'articulació, que en algunes varietats s'expressa d'una manera i en altres, d'una altra; la major freqüència de certes paraules, que representa una ruta més curta per a aquestes paraules pel que fa a les xarxes neuronals del nostre sistema nerviós.

Cal distingir entre *evolució* i *història*. La història és el procés d'analitzar el sistema com un producte de canvis evolutius irreversibles i propietats emergents en cada estat del sistema (Ingold, 2000; Tarride, 1995).

En la descripció dels sistemes dinàmics sempre hi ha un element històric. Els sistemes dinàmics no són estables en un moment donat (sí que són *meta-estables*; consulteu el subapartat 1).

En biologia res no té sentit sense evolució (al contrari del que succeeix en altres camps; per exemple, no cal comprendre l'origen dels ordinadors per saber-ne utilitzar un o per beneficiar-se'n). I així és en la lingüística.

La llengua considerada com un CAS ens permet comprendre el comportament general que sorgeix de les interaccions locals entre un gran nombre d'agents. Molt sovint aquest comportament general (i la dinàmica emergent) és complex, però no està determinat per cap disseny previ ni està subjecte a un mecanisme de control centralitzat. És difícil predir quines seran les propietats generals emergents del sistema basant-nos només en el coneixement de les parts que el conformen.

Com sabem, el canvi és un element essencial en el comportament dels sistemes complexos. Els sistemes són adaptables en el sentit que tenen una capacitat per evolucionar en resposta a un entorn canviant. De les interaccions locals, en poden sorgir models globals complexos amb noves propietats.

En cibernètica els sistemes oberts són conjunts que interactuen amb entrades, rendiments i sortides d'energia i informació. A través de la retroalimentació negativa, els sistemes romanen en equilibri dinàmic. Els conceptes essencials en el model de la cibernètica són: *codificació* (activadors de comunicació), *entropia negativa* (organització de l'energia de l'entorn per mantenir el sistema) i *equifinalitat* (diferents vies de desenvolupament poden dur a la mateixa destinació). Tots aquests conceptes els considero capitals en lingüística.

El model bàsic de la cibernètica presenta un sistema controlat, retroalimentació i tres components: detector (informació), selector (preferències) i efector (acció). El comportament individual es pot analitzar amb aquesta formulació. Per exemple, en el comportament interpersonal, el detector és la comunicació, la transacció és el selector i l'organització és l'efector. Els components interpersonals poden ser combinats amb els components socials. Kuhn (1975) emprava una estructura conceptual que considera que les disciplines bàsiques de les ciències socials no són la sociologia, l'economia i la ciència política, sinó la comunicació, la transacció i l'organització. La llengua és l'eina bàsica per al funcionament dels sistemes individuals, interpersonals i socials.

4.2. Procés d'estandardització de les llengües

Des del meu punt de vista, en aquesta àrea de la cibernètica és interessant establir un paralelisme entre com, d'una banda, un canvi es propaga a través d'una societat —sempre d'acord amb un procés gradual tant dins la llengua,

a través del lèxic, com fora, en els diferents estrats socials i regions geogràfiques— i com, d'altra banda, ocorren els processos d'estandardització (Massip, 1994).

1) Pas 1: procés de selecció

Coşeriu (1992) va definir la selecció en el canvi lingüístic com l'ús alternatiu d'una forma antiga i una nova tradició, de manera que qualsevol canvi és la selecció d'una forma i l'exclusió d'altres.

En els processos d'estandardització sempre hi ha un pas de selecció: primer, de la varietat (o varietats, en els tipus policèntrics d'estandardització) i després, dels elements (d'un conjunt de formes històriques).

La selecció és sempre el resultat de tensions dinàmiques que determinen el triomf d'una solució, i aquest triomf té la base en solucions adoptades en altres parts del sistema, per això és difícil predir la forma lingüística que es generalitzarà.

2) Pas 2: procés de generalització

Hi ha dos tipus de generalització: extensiva i intensiva.

Es produeix una generalització extensiva entre tots els parlants d'un grup, mentre que la generalització intensiva apareix en totes les paraules que contenen el fonema afectat o una altra classe paradigmàtica.

Pel que fa a la norma, el concepte de generalització extensiva és força important. Es pot dir que el canvi lingüístic s'ha generalitzat quan els parlants fan servir les formes seleccionades de manera espontània, sempre o de vegades alternant amb les formes antigues. Podem dir que una norma s'ha generalitzat quan s'accepta i s'utilitza una forma sense que els usuaris de la llengua la qüestionin.

La distinció entre desviació i pertorbació és molt interessant quan es tracten l'estandardització i altres temes sociolingüístics que s'hi relacionen.

- *Desviament*: canvis en la situació d'un agent, sobre els quals l'agent no té control (encara que l'agent pot mirar de controlar-ne els efectes).
- *Pertorbació*: desviació negativa. Les pertorbacions poden originar-se en el medi o poden sorgir del mal funcionament d'un agent.

4.3. Cognició

- **Cognició:** adquisició, processament, emmagatzematge i ús de la informació i del coneixement (vegeu *coneixement*, a continuació) per donar suport a la presa de decisions intel·ligents.

Una perspectiva que s'ha de tenir en compte quan parlem de lingüística i canvi lingüístic és la de la cognició col·lectiva (o social) (Heylighen, 2013). Per a les llengües és una perspectiva molt rellevant, ja que significa fonaments cognitius innats i adquirits. I també perquè, per tal que els canvis es converteixin en part de sistema, sempre han de generalitzar-se en la societat.

El cervell social implica una xarxa de regions corticals i subcorticals que són especialment sensibles a les situacions socials. Cada membre d'una comunitat, amb les seves característiques cognitives i el seu cervell (i el seu cervell social), participa en la interacció social i lingüística i contribueix a conformar la societat a la qual pertany. Les contribucions dels estudis sobre el cervell social estan canviant la manera d'entendre els motius i les característiques de molts fenòmens socials, i són absolutament necessàries per explicar la societat contemporània en tots els seus aspectes, inclòs el lingüístic. El paper del cervell social en la configuració de la llengua era clar fins i tot abans que es descobrís el concepte de *cervell social*.

La cognició social requereix habilitats com ara el reconeixement de l'afecte, la memòria i el reconeixement facial, una interpretació adequada de l'afecte i la prosòdia, i el coneixement de les ments dels altres (com demostra la teoria de la ment).

- **Varietat d'accions:** quantitat d'accions que un agent pot executar potencialment. Com més gran sigui la varietat d'accions d'un agent, més gran serà la varietat de desviacions amb les quals l'agent pot lluitar, perquè els diferents tipus de desviacions generalment requereixen diferents tipus d'accions. Aquesta és una generalització de la coneguda «lei de la varietat requerida» d'Ashby (1976).

La distinció entre informació i coneixement és significativa: la informació és qualsevol canvi de condició que redueix la incertesa de l'agent, mentre que el coneixement és la capacitat derivada de l'experiència, l'educació i/o la comunicació per anticipar les conseqüències d'un estat o una acció determinada. El coneixement difereix de la informació en el fet que produeix prediccions generals que són aplicables a diferents situacions, mentre que la informació, estrictament parlant, només s'aplica a la situació actual.

- *Coneixement*: capacitat, típicament derivada de l'experiència, de l'educació o de la comunicació, per anticipar les conseqüències d'un estat o una acció determinada.

- *Intelligència*: donada certa informació, grau en què un agent pot prendre bones decisions, és a dir, selecció d'accions que acumulen la màxima utilitat a llarg termini. La intelligència té dos components: 1) coneixement (vegeu el paràgraf anterior) i 2) intelligència fluida (la capacitat d'explorar internament moltes combinacions diferents de possibles esdeveniments i accions per trobar la que, segons el coneixement existent, tindrà major utilitat).

En aquesta secció s'ha d'esmentar que el funcionament del cervell ens ajuda a emmarcar la comprensió de com construïm el significat (i les xarxes semàntiques).

El significat d'un concepte depèn de la manera en què està connectat a altres conceptes. Hi ha molts tipus de relacions semàntiques (entre paraules, entre significat i significat, entre paraules del mateix origen i família lèxica, en terminologies específiques, etc.), i tot això, deixant a banda de la complexitat dels significats en si mateixos.

Depenent del tipus de relació que s'estudii, presentarem una estructura de xarxa específica que ens permetrà veure les associacions entre els conceptes.

El model de funcionament cerebral és clau per ajudar-nos a comprendre com es produeixen totes aquestes relacions. És per això que avui en dia els neurocientífics estan treballant amb xarxes neuronals artificials.

Les xarxes neuronals són algoritmes d'intelligència artificial que imiten com funcionen les xarxes bioelèctriques en el cervell (formades per les neurones i les seves sinapsis). Simulen l'arquitectura i el comportament del cervell humà i del sistema nerviós.

La memòria també funciona com una xarxa. El cervell humà vincula la informació que prové de l'exterior amb altra informació que ja existeix en la memòria. Una vegada el cervell detecta les pistes d'associació, busca dades emmagatzemades que coincideixin amb les pistes d'associació de la nova informació i, finalment, elimina els elements que no considera útils i emmagatzema en la memòria el que ha cregut que podria ser útil en el futur.

El coneixement de qualsevol tipus s'emmagatzema en la connexió i en les correlacions entre les unitats de processament, no en les mateixes unitats. Aquestes unitats serien equivalents a les neurones o xarxes de neurones i les connexions serien les sinapsis. El coneixement lingüístic semàntic també es troba en les relacions, no en paraules individualitzades.

En l'adquisició d'una segona llengua també és molt útil considerar el coneixement i la intelligència com a habilitats diferents que s'han d'adquirir per ser

competent en una llengua apresada (a més, òbviamment, de les característiques cognitives dels agents).

4.4. Interaccions

- *Interacció*: efecte recíproc de dos (o més) agents (per exemple, A i B) entre si. Per exemple, l'acció realitzada per A crea una condició que desencadena una altra acció (reacció) de B. Aquesta segona acció, al seu torn, afecta la condició de A i l'estimula a reaccionar, i així successivament.

- *Sinergia*: qualsevol guany causat per una interacció —una interacció amb una suma positiva—. L'oposat a la sinergia és la fricció, i la relació entre els agents involucrats en una interacció sinèrgica o de suma positiva és la cooperació.

La relació entre els agents involucrats en una interacció amb fricció o una suma negativa és de *conflicte*.

Tots aquests termes estan en ús actualment en la terminologia sociolingüística. Podem veure que generalment són aplicables a qualsevol interacció en sistemes complexos.

Aquests conceptes també es poden aplicar a la conversa, de manera que a partir de la participació lingüística interactiva de dos o més actors es pot generar *sinergia* i, per tant, creació de coneixement, comprensió i cooperació. O *fricció*, en la qual la interacció lingüística serveix només per concloure el conflicte. En qualsevol cas, la interacció lingüística (encara que no només la llengua oral) és gairebé l'única manera d'aconseguir la cooperació.

La sinergia en la teoria de sistemes és el procés entre les forces antitètiques de la natura que porta a un desenvolupament d'estructures fortament complexes (Parra, 1992: 437-443).

4.5. Sistemes

- *Sistema*: grup d'agents units per una restricció compartida o una xarxa d'enllaços. Els agents en el sistema poden ser vistos com a components del sistema.

Quan els agents en un sistema comparteixen un objectiu, el sistema funciona com un agent d'ordre superior.

- *Supersistema*: sistema els components del qual són altres sistemes, anomenats *subsistemes*.

- *Sistema complex adaptatiu*: sistema que consta de molts agents interactius, on les seves interaccions no són rígidament fixes, preprogramades o controlades, sinó que s'adapten contínuament als canvis en el sistema i en el seu entorn.

Els problemes als quals actualment s'enfronten els individus, les organitzacions i la societat en general es refereixen a sistemes complexos i evolutius, com l'ecosistema global, la societat, l'economia i el nostre propi sistema intern de pensaments i emocions (vegeu el subapartat 3).

- *Cognició distribuïda*: adquisició, emmagatzematge i ús de la informació i el coneixement distribuïts entre diferents agents en un sistema, per tal de donar suport a la seva intel·ligència col·lectiva (Heylighen, Heath i Van Overwalle, 2004).

La llengua és un dels exemples d'aquests mecanismes de suport per a l'intercanvi d'informació.

Molts fenòmens emergents que observem en sistemes adaptatius complexos es poden entendre com a resultat de canvis en l'equilibri i l'estabilitat, quan el sistema es mou entre equilibris. Aquests canvis poden ser el resultat d'un canvi gradual o sobtat i en general es denominen transicions crítiques quan el sistema arriba als punts d'inflexió.

Quan els sistemes complexos arriben (o són a prop) dels atractors (vegeu el subapartat 4.1 *meta*), demostren estabilitat i variabilitat.

Quan un sistema o una de les seves parts canvia d'un atractor relativament estable a un altre en un procés de *transició de fase*, el punt de transició es pot marcar perquè augmenta la variabilitat del comportament. Quan els sistemes passen per una transició de fase, s'autoorganitzen i l'organització que emergeix pot ser nova i pot diferir qualitativament de les diferents organitzacions dels sistemes o de les seves parts (Massip, 2013). L'estat natural d'un sistema lingüístic pot definir-se com una adaptació dinàmica a un context específic.

- *Restricció*: limitació en la varietat d'accions per a alguns agents.

Un *atractor* (vegeu també el subapartat 4.1 *meta*) és una restricció que evolucionaria espontàniament.

- *Enllaç o connexió*: interacció estabilitzada entre dos agents que exerceix una restricció en la seva acció posterior.

- *Ambient*: tot el que es considera extern a un agent o un sistema determinat, però encara interactua amb ell. Això vol dir que l'entorn proporciona les condicions inicials o l'entrada que desencadena una interacció, al mateix temps que accepta la sortida o el canvi produït per aquesta acció.

En lingüística, hem de parlar de l'entorn dels parlants i de l'entorn de les paraules dins el propi sistema lingüístic (que comunament anomenem context).

- *Mitjà*: entorn compartit pels agents que interactuen, però també pot ser intern als agents.

4.6. Organització

- *Organització*: xarxa estabilitzada d'interaccions entre agents que funciona per garantir la coordinació de la seva acció i especifica els rols i les interaccions particulars entre els agents del sistema. Es pot imposar des de l'exterior o pot sorgir de l'autoorganització.

- *Coordinació*: entre els agents, agrupa els seus recursos, tant els materials com els informatius, perquè el grup en el seu conjunt pugui actuar de manera més efectiva i intel·ligent que cada agent individualment. Aquest és l'origen de la intel·ligència col·lectiva. A més, el sistema augmenta encara més el seu coneixement i la seva intel·ligència perquè interactua amb el seu entorn, alhora que aprèn d'aquestes interaccions perquè sigui cada vegada més efectiu a anticipar fenòmens i triar les accions apropiades. Així, la xarxa de relacions entre els agents comença a comportar-se com una xarxa neuronal, capaç de dur a terme processos cognitius cada vegada més sofisticats. Aquests sistemes autoorganitzats es distribueixen en la seva organització: en general és difícil distingir components o subsistemes separats que fan funcions separades; els components cooperen en conjunt.

- *Mediador*: estructura reguladora externa als agents que promou la coordinació entre ells. Pot sorgir de l'autoorganització o pot ser imposat per un agent intern o extern.

- *Estigmèrgia*: forma de coordinació indirecta a través del mitjà, on la traça deixada per una acció en el medi estimula l'acompliment d'una acció posterior, cosa que permet continuar amb el treball que ja s'ha fet (Heylighen, 2007ab; Van Dyke Parunak, 2006).

Aplicat a les llengües, aquest concepte és útil en sociolingüística: segons l'entorn i la llengua amb la qual una persona s'acosta a una altra (selecció influenciada pel prestigi o per altres raons), es generen uns tipus d'usos o uns altres entre els que intervenen en el procés lingüístic.

5. CAUSES PROXIMALS I DISTALS

Les causes del canvi són mecanismes generadors de canvi en els sistemes. I també són les narratives que permeten als agents humans expressar aquests canvis de manera significativa.

Les causes *proximals* afecten el fenotip (morfologia) i el comportament, mentre que les causes *distals* afecten el genotip i la seva història. L'experimentació típicament ajuda en la determinació de les causes proximals, mentre que les causes distals o finals es determinen per inferència basada en narracions històriques (Rovelli, 2016: 135). Gairebé tots els problemes ecològics involucren causes proximals i distals (i això també és cert per a la lingüística), perquè l'ecologia (com la lingüística) tracta amb sistemes altament complexos.

La relativitat general ens ensenya que l'espai no és una caixa inerta, sinó que és dinàmic: una mena de closca de cargol mòbil en el qual estem continguts, un cargol que es pot comprimir i torçar. Jo diria que, metafòricament, els canvis d'espai amb els paràmetres socials, polítics, econòmics o d'altres tipus canvien.

L'altra conseqüència de la teoria de la relativitat és aquesta: de la mateixa manera que la idea d'un espai continu que conté coses desapareix, també ho fa la idea d'un temps elemental i primari que flueix independentment de les coses. Això vol dir que el canvi, en el temps i en l'espai, és ubic.

6. L'OBSERVADOR

L'atenció a l'observador comença mà per mà amb el naixement de la filosofia a l'antiga Grècia, continua amb la concepció del subjecte i l'objectivitat, el desenvolupament de la ciència mecanicista i el positivisme en el món modern, i arriba a la revolució de la relativitat, al constructivisme i al pensament complex avui.

La complexitat inclou l'observador en el cor de la seva teoria.

Vallée (1990: 240) fa servir tres formes diferents per referir-se a la complexitat: el «sentiment de l'observador», el «judici de l'observador» i el «nom» d'una relació subjecte-objecte (una expressió objectiva i una de subjectiva, que la relació no es posa en confrontació). Ell veu la realitat com una construcció mental.

Segons Morin (1998), el mètode per actuar sobre sistemes complexos no pot ser el mètode científic (o el que avui es diu científic, un terme el significat del qual està experimentant un procés d'evolució) i demana que s'abordi la complexitat a través d'un procés d'acció i de reflexió.

Quines eines o instruments de reflexió i acció tenim? Educació, cultura, ciència, tecnologia, democràcia, justícia.

Bachelard (1985: 130) assenyala que tot fenomen de la realitat és una xarxa de relacions. Les idees simples són hipòtesis de treball, conceptes de treball, que s'han de revisar per assignar-los el paper epistemològic més adequat.

La nova cibernètica (segle XXI) considera que la informació està construïda i reconstruïda per un individu que interactua amb l'entorn. En considerar la ciència com a dependent de l'observador, es proporciona una nova base epistemològica a la ciència.

8. CONSIDERACIONS FINALS

Podríem visualitzar els processos que estan produint canvis en una llengua en un moment donat com un conjunt d'engranatges en què les rodes dentades estan girant, algunes d'acord amb les lleis regulars de l'evolució i d'altres segons els processos d'evolució irregular. Al mateix temps, cada roda dentada té diferents velocitats de moviment.

Pel que fa a les llengües romàniques, per exemple, una d'aquestes rodes dentades inclouria elements lèxics basats en el llatí, altres inclourien elements provinents d'altres llengües, i altres inclourien elements de llengües especials. Vist d'aquesta manera, es fa més clar que cada roda dentada es mou a una velocitat diferent. El conjunt d'engranatges segueix girant a mesura que passa el temps i es produeixen emergències a partir de la seva interacció amb l'entorn i de la dels elements entre si (fins i tot entre els elements que hem descrit com a inclosos en diferents rodes dentades).

Així doncs, les paraules patrimonials no segueixen les mateixes regles d'evolució fonètica que els cultismes (per exemple, de l'ètim RADĪU tenim en català la paraula patrimonial *raig* i el semicultisme *ràdio*); les paraules importades d'altres llengües tampoc no segueixen les regles del seu propi canvi (per exemple, *merci* català [mɛrsi] del *merci* francès, molt comú avui en dia en el català oriental, mentre que *mercès* és una forma més antiga per dir gràcies i d'acord amb les regles d'evolució del català). Les paraules patrimonials sovint tenen adaptacions dels fonemes que són típics de l'inventari de la seva pròpia llengua

(com en el cas del català *maco*, on el so espanyol [χ] (de *majo*), que no està en el sistema fonològic del català, va ser adaptat a [k]).

En conclusió, el canvi lingüístic sorgeix com a conseqüència de l'aparició constant de fenòmens variacionals resultants de les interaccions dels individus entre si i amb l'entorn, i també de les interaccions entre els elements del sistema lingüístic.

Després de considerar el canvi i el vector temps en lingüística, proposo d'adaptar semànticament el terme *diacronia* i alterar-ne la intenció, de manera que la diacronia es refereixi a l'estudi integrador de la variació lingüística, la veritable essència de la lingüística, que sempre té una dimensió temporal (Massip, 2003 i 2007a). Parlem de sincronia dinàmica i això és diacronia. No veig la possibilitat de fer diacronia sense integrar la sincronia i, per descomptat, no és possible fer sincronia sense diacronia, ja que la diacronia està ja integrada en la sincronia.

El nostre següent repte és aplicar aquest marc a una llengua específica i utilitzar-lo quan tractem altres temes lingüístics.

REFERÈNCIES BIBLIOGRÀFIQUES

- ANDRASON, A. (2014). *A complex system of complex predicates: Tense, taxis, aspect and mood in base mandinka from a grammaticalization and cognitive perspective*. Tesi doctoral supervisada per la Dra. Marinna Visser, Stellenbosch University.
- ASHBY, W. R. (1976). *Introducción a la cibernética*. Buenos Aires: Nueva Visión.
- BACHELARD, G. (1985). *O novo espírito científico*. Rio de Janeiro: Tempo Brasileiro.
- BECKNER, C.; BLYTHE, R.; BYBEE, J.; CHRISTIANSEN, M. H.; CROFT, W.; ELLIS, N. C.; HOLLAND, J.; KE, J.; LARSEN-FREEMAN, D.; SCHOENEMANN, T. (2009). «Language is a complex adaptive system: position paper». *Language Learning*, núm. 59, pàg. 1-26.
- COŞERIU, E. (1992). «Le changement linguistique n'existe pas». *Communication et Cognition: An Interdisciplinary Quarterly Journal*, núm. 25, pàg. 121-135.
- HEYLIGHEN, F.; HEATH, M.; Van OVERWALL, F. (2004). *The emergence of distributed cognition: A conceptual framework*. www.researchgate.net/publication/249812898_The_Emergence_of_Distributed_Cognition_a_conceptual_framework.
- HEYLIGHEN, F.; CILLIERS, P.; GERSHENSON (2007). *Complexity and philosophy: Complexity science and society*. Oxford: Radcliffe Publishing.
- HEYLIGHEN, F. (2013). «Self-organization in communicating groups: the emergence of coordination, shared references and collective intelligence». A: MASSIP-BONET, A.; BASTARDAS-BOADA, A. (eds.). *Complexity perspectives on language, communication and society*. Berlín / Heidelberg: Springer, pàg. 117-149.

- INGOLD, Tim (2000). *The perception of the environment: Essays in livelihood, dwelling and skill*. Londres: Routledge.
- KUHN, A. (1975). *Unified social science: A system-based introduction*. Illinois: Dorsey Press.
- LEE, N.; MIKESELL, L.; JOACQUIN, A. D.; MALTES, A. W.; SCHUMANN, J. H. (2009). *The interactional instinct: The evolution and acquisition of language*. Oxford: Oxford University Press.
- MASSIP, À. (1994). «Observations sur la variation linguistique: des exemples dans la langue catalane médiévale». *Communication & Cognition*, vol. 27, núm. 1/2, pàg. 337-349.
- MASSIP, À. (1997). «La variation linguistique: perspectives apportées par l'étude intégrationnelle de la langue». *Cahier d'Études Romanes*, núm. 9, pàg. 114-123.
- MASSIP, À. (2003). «Canvi i variació: de la dinamicitat de la llengua a la lingüística integrativa». A: *Miscel·lània Joan Veny*. Vol. 2. Barcelona: Publicacions de l'Abadia de Montserrat, pàg. 5-18.
- MASSIP, À. (2007a). «Canvi i variació: Noves metàfores, noves prospeccions». A: *Actes del Congr s de l'AILLC (Universitat de Girona 8-13 de setembre de 2003)*. Barcelona: Publicacions de l'Abadia de Montserrat, pàg. 235-243.
- MASSIP, À. (2007b). «La variabilit  en langue: Th orie et complexit ». *Phrasis: Studies in Language and Literature*, vol. 46, n m. 2, p g. 85-99.
- MASSIP-BONET, À. (2013). «Language as a complex adaptive system: towards an integrative linguistics». A: MASSIP-BONET, À.; BASTARDAS-BOADA, A. (2013). *Complexity perspectives on language communication and society*. Berl n / Heidelberg: Springer, p g. 35-60. www.springer.com/engineering/robotics/book/978-3-642-32816-9.
- MASSIP-BONET, À. (2019). «Linguistic variation and change: approach from the perspective of complex adaptive systems». A: MASSIP-BONET, À.; BEL-ENGUIG, G.; BASTARDAS-BOADA, A. (2019). *Complexity applications in language and communication sciences*. Switzerland: Springer Nature, p g. 77-92.
- MORIN, X. E.; MORIN, E. (1977). *La m thode*. Par s: Seuil.
- MORIN, E. (1998). *El m todo 3: el conocimiento del conocimiento*. Madrid: C tedra.
- MUFWENE, S. (2013). «The emergence of complexity in language: An evolutionary perspective». A: MASSIP-BONET, À.; BASTARDAS-BOADA, A. (eds.). *Complexity perspectives on language, communication and society*. Berl n / Heidelberg: Springer, p g. 197-218.
- MUKHERJEE, S. (2017). *El Gen. Una historia  ntima*. Barcelona: La Campana.
- PARRA LUNA, F. (1992). *Elementos para una teor a formal del sistema social*. Madrid: Complutense.
- ROVELLI, C. (2016). *Set lli ons breus de f sica*. Barcelona: Anagrama.
- TARRIDE, M. T. (1995). «Complejidad y sistemas complejos». *Hist ria, Ci ncias, Sa de-Manguinhos*, vol. II, n m. 1, p g. 46-66.

- TURCHIN, V. F. (1993). «The cybernetic ontology of action». *Kybernets*, vol. 22, núm. 2, pàg. 10-30.
- VALLÉE, R. (1990). «Sur la complexité d'un système relativement à un observateur». *Revue Internationale de Systémique*, vol. 4, núm. 2, pàg. 239-43.
- VAN DYKE PARUNAK, H. (2006). «A survey of environments and mechanisms for human-human stigmergy». A: WENYNS, D.; VAN DYKE PARUNAK, H.; MICHEL, F. (eds.). *Environments of multi-agent systems II*, Vol. 3830 of Lecture Notes in Computer Science (LNCS), pàg. 163-186.

Mètodes de processament per a l'anàlisi lingüística de la variació

Irene Castellón Masalles
Universitat de Barcelona

1. INTRODUCCIÓ

La lingüística computacional i el processament del llenguatge natural (PLN) han patit molts canvis en aquest segle XXI. Internet i l'aparició de les xarxes socials han propiciat que es pugui disposar de textos en moltes llengües, de diferent registre i de gran volum.

Aquest fet ha revolucionat la recerca en aquesta àrea, en el sentit que l'objectiu és el tractament del llenguatge general enfront del tractament de dominis concrets que era el que predominava en el segle XX. D'altra banda, la disponibilitat de grans volums de dades textuais ha permès el desenvolupament i l'aplicació de tècniques d'aprenentatge automàtic enfront dels sistemes basats en regles que requereixen molt d'esforç per a la seva construcció. Així, la recerca en PLN s'obre al tractament estadístic del llenguatge, tot prenent idees i mesures de la teoria de la informació i dels teoremes matemàtics (Manning i Shultze, 1999).

Aquestes tècniques poden ser utilitzades com a instruments per a l'estudi lingüístic dels textos, tant per demostrar hipòtesis com per capturar patrons complexos, tal com usem la calculadora per fer càlculs.

Podem parlar de dues tècniques. Els sistemes d'anàlisi lingüística ens permeten etiquetar textos de manera automàtica molt ràpidament, fet que permet disposar de moltes dades anotades. L'índex d'error d'aquestes anotacions depèn del nivell d'anàlisi lingüística; aquesta tècnica és la que obté millors resultats en la morfologia i la semàntica és la que presenta més errors.

L'altra línia és la dels sistemes inductius actuals, que són capaços de construir models de llenguatge només a partir de les dades. En realitat, per al desenvolupament d'aquests sistemes calen corpus; en alguns casos els corpus poden ser col·leccions de textos sense anotar, que anomenem text pla, o bé textos anotats, on la informació s'ha fet explícita. Aquests sistemes es poden aplicar en la recerca lingüística, on proporcionen dades interessants (vegeu els apartats 4 i 5 del capítol de Bel-Enguix en aquest llibre). En aquest article presen-

tem alguns d'aquests mètodes aplicats sobre un corpus del català dividit en diverses particions diacròniques, per tal de veure la variació en els diferents períodes.

L'objectiu d'aquest treball és presentar una metodologia aplicada als corpus històrics, basada en l'aplicació de mesures i eines utilitzades en PLN, per a l'estudi de la variació diacrònica. Aquests diferents mètodes exploren els corpus de manera quantitativa, van més enllà del tractament formal de la paraula i permeten d'explorar aspectes semàntics del vocabulari.

2. EL CORPUS

Per a l'anàlisi lingüística de la variació aplicada a la diacronia utilitzarem com a banc de dades diversos corpus: el Corpus Informatitzat del Català Antic (CICA) (Torruella *et al.*, 2009), el Corpus Textual Informatitzat de la Llengua Catalana (CTILC) i el corpus Scripta (Veny i Massip, 2009).

El CICA és un corpus històric del català que recull textos des del segle XI al XVIII. Presenta els textos sense anotació i té un total de 8.689.126 ocurrències.

El corpus Scripta és molt més petit, té unes 870.000 ocurrències escrites, des del segle XIII al XXI, però té la característica que inclou textos no habituals.

El CTILC és un corpus del català escrit del període 1830-2003, corresponent a una mica menys de dos segles, i està lematitzat i categoritzat en relació amb la morfologia.

Es pot observar en la figura 1 que el CTILC i el corpus Scripta van ampliant el volum de paraules com més a prop estan de l'actualitat. Ben al contrari, el corpus CICA té com a segles més representatius del XIII al XVI.

El corpus més gran i més anotat, el CTILC, conté seqüències de lemes + categories (vegeu la figura 2). Això implica que se li han extret les paraules funcionals (determinants, preposicions i conjuncions), de manera que es compon de la llista de lemes i categories emprades en cada frase. El CTILC inclou, per tant, només unitats lèxiques, un total de 33.552.815 ocurrències.

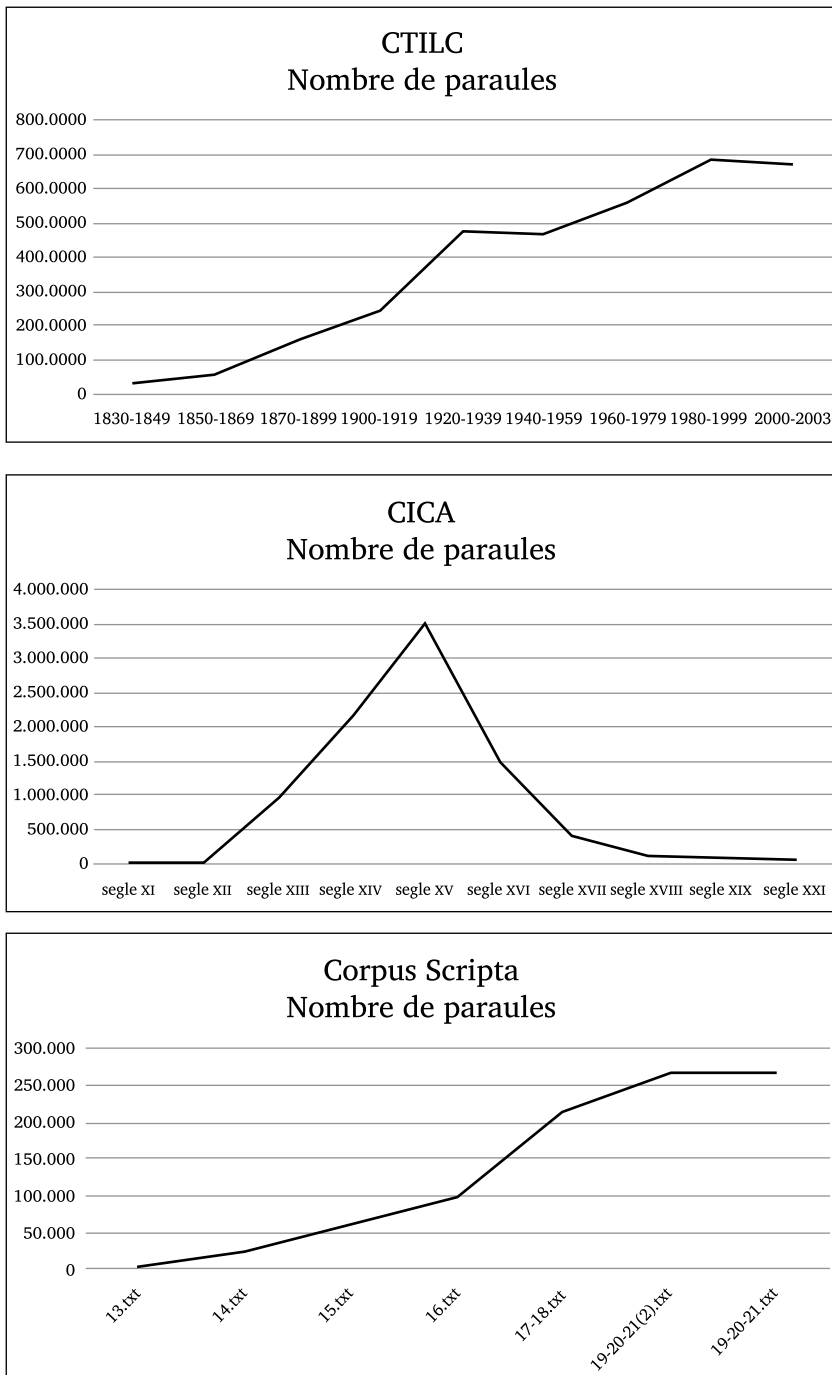


Figura 1. Representació per anys o segles del corpus analitzats.

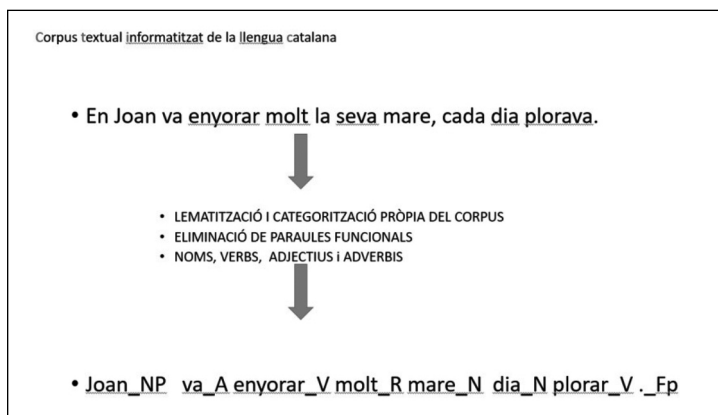


Figura 2. Exemple d'una frase en el corpus CTILC.

Per tal de treballar amb tots els corpus com si fos un de sol és necessària l'harmonització d'aquests corpus, és a dir, que tots presentin el mateix tipus de dades. Però el CICA i l'Scripta estan constituïts per formes, mentre que el CTILC està constituït per lemes, per la qual cosa no ha estat possible treballar amb tots els corpus com si fossin un de sol. Això sí, podem utilitzar-los comparativament si només estudiem un conjunt de termes. Per il·lustrar aquest fet, observem en la figura 3 l'ús de la paraula *jaure* i la variant *jeure* en dos corpus (CICA i CTILC). Hem agrupat les diferents formes del CICA i les comparem amb els resultats del CTILC. Com podem veure i calia esperar, els dos corpus són complementaris pel que fa a les diferents franges temporals.

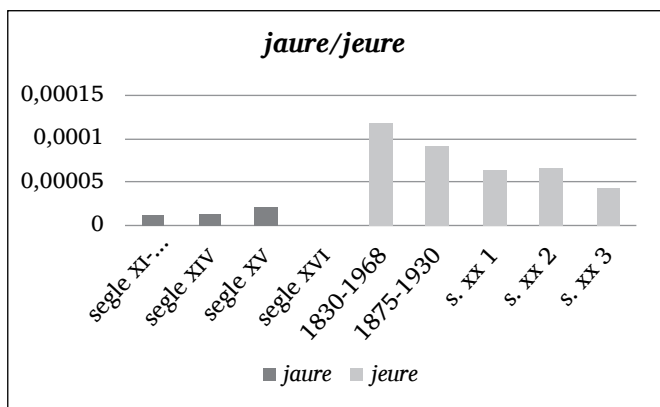


Figura 3. L'ús de *jaure* i *jeure* en els dos corpus CICA i CTILC.

Com hem dit, l'objectiu d'aquesta recerca és establir una metodologia per a l'estudi de la variació lingüística al llarg del temps, per això ha estat necessari dividir els corpus en diferents particions segons l'any de la seva creació. Exemplificarem aquest procés amb el CTILC.

No és una tasca simple fer la divisió per anys, ja que els resultats poden variar segons com es divideix el corpus. Per això hem fet diverses particions, així es poden comparar els resultats i entendre les diferències. S'han fet tres tipus de particions del corpus de l'IEC (CTILC): una per a cada canvi de règim polític (partició 1), aproximadament una cada cinquanta anys (partició 2) i aproximadament una cada vint anys (partició 3) (vegeu la figura 4).

Partició 1	Partició 2	Partició 3
1830-1868	1830-1869	1830-1849
1869-1874	1870-1899	1850-1869
1875-1930	1900-1959	1870-1899
1931-1939	1960-1999	1900-1919
1940-1975	2000-2003	1920-1939
1976-2003		1940-1959
		1960-1979
		1980-1999
		2000-2003

Figura 4. Particions fetes en el CTILC.

Si ens fixem en la primera partició (taula 1), la que s'associa als canvis de règim, el *nombre* d'ocurrències varia molt entre períodes, els més antics tenen menys ocurrències i els més curts també en tenen menys que la resta.

Aquestes diferències poden superar-se amb certes mesures que ja tenen en compte el volum total del corpus, però són diferències que sempre haurem de considerar en l'anàlisi dels resultats.

Partició	Ocurrences totals
1830-1868 (isabelí)	836.857
1869-1874 (Sexenni Liberal)	334.155
1875-1930 (Restauració borbònica)	6.470.870
1931-1939 (República)	2.888.576
1940-1975 (franquisme)	8.186.055
1976-2003 (Règim del 78)	5.378.649

Taula 1. Primera partició basada en règims històrics i el volum de paraules que contenen.

3. MESURES ESTADÍSTIQUES

Comptant paraules podem arribar a caracteritzar certs aspectes dels textos; utilitzant diverses mesures estadístiques podem anar una mica més enllà. Segons l'objectiu de la recerca, aquests recomptes poden aplicar-se a textos no anotats, així com a textos anotats. Algunes de les mesures més habituals o que tenen una major explotació per a la comparació de textos que aquí tractarem són la densitat lèxica, que mesura la riquesa lèxica dels textos, $tf*idf$, que mesura la rellevància d'una paraula en un document, i la informació mútua, que mesura la coocurrència d'un parell de paraules.

La densitat lèxica consisteix a posar en relació les ocurrences de les formes (*tokens*) amb la varietat de formes (*types*). El càlcul es fa amb l'equació de (a).

(a) Densitat lèxica = nombre de *types* / nombre d'ocurrences

El resultat, sempre normalitzat entre 0 i 1, permet assignar un índex de densitat lèxica a textos, de manera que podem comparar la riquesa lèxica de diferents textos més objectivament.

En el CICA la densitat lèxica que obtenim és baixa, ja que conté 8 milions d'ocurrences i uns 522.035 *types*. Una de les causes pot ser la diversitat de variants a l'hora d'escriure la paraula en el català antic. En l'*Scripta* la densitat lèxica és més alta, segurament perquè inclou més quantitat de textos normativitzats (870.000 ocurrences i 18.778 *types*) i perquè és molt més petit, concretament, un 10% del CICA (vegeu la figura 1).

Per tant, aquesta mesura s'ha d'interpretar tenint en compte les característiques del corpus i és totalment comparable quan els corpus són d'una època similar quant a la normativització i, a més, la mesura dels corpus no és gaire diferent.

Per exemple, prenem com a corpus les particions segmentades del CTILC en nou parts. Cadascuna comprèn vint anys, excepte la darrera, que és una mica més petita. En la figura 5 comparem la densitat lèxica de cadascuna, en el primer cas sense equilibrar les particions i en el segon amb les particions equilibrades quant al nombre de paraules.

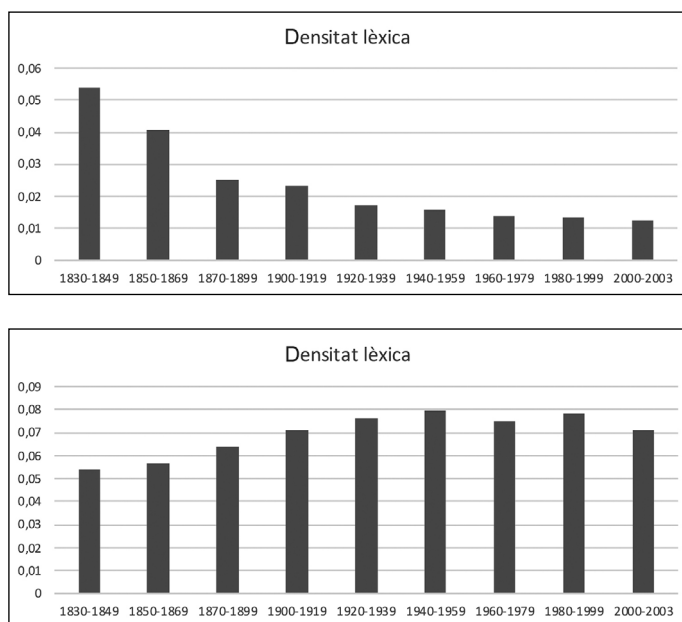


Figura 5. Densitat lèxica calculada per a cada partició, a l'esquerra sense equilibrar el corpus, a la dreta els resultats de la densitat lèxica amb les particions equilibrades.

Per tant, aquesta mesura decreix quan els corpus són grans. Llavors és recomanable calcular aquesta mesura en corpus equilibrats, és a dir, en corpus amb més o menys la mateixa extensió. Si iguaem les particions i agafem només una quantitat determinada de cadascuna, la densitat lèxica de les particions es mou entre 0,5 i 0,8. Aquesta mesura, però, està condicionada pel fet que s'ha calculat només amb les unitats lèxiques i no s'han tractat les dades amb paraules funcionals, i per tant no és comparable amb els altres corpus que són textuals.

La informació mútua (IM) (Manning i Schutze, 1999) capta el fet que una paraula aparegui en el mateix context d'una altra amb una certa freqüència, és a dir, ens proporciona un índex objectiu sobre l'atracció de determinades paraules en les produccions. La formula és la indicada a (c).

$$(c) I(x_i; y_j) = \log P(x_i | y_j) / P(x_i)$$

La informació mútua posa en relació la probabilitat condicionada de l'aparició de dues paraules amb la probabilitat d'aparició d'una d'elles en tot el text. Així, la informació mútua de dues variables aleatòries mesura la dependència mútua entre dues variables. Per tant, la IM ens proporciona una mesura per captar la coocurrència dels mots en els diferents contextos.

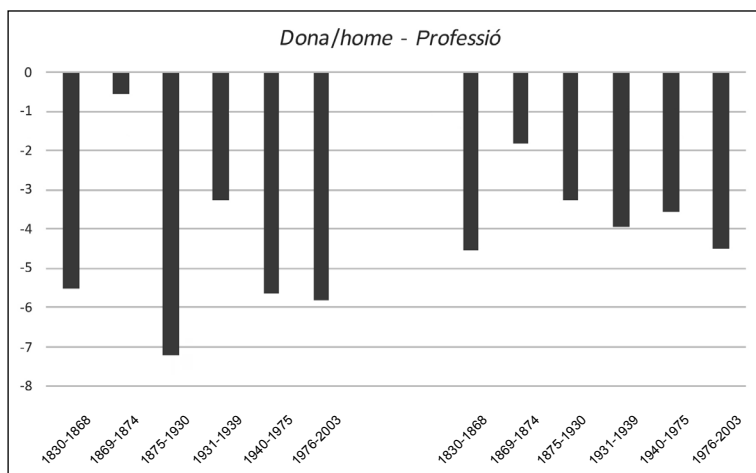


Figura 6. Informació mútua de *dona* i *professió* (esquerra) i d'*home* i *professió* (dreta) en les particions segons el règim polític.

Com a exemple, hem calculat la IM de les paraules *dona* i *professió* en la primera partició, on s'identifica cada període amb un règim polític. En aquest gràfic (figura 6), les línies més baixes indiquen que en la partició corresponent hi ha menys representativitat de la coocurrència. Hem de tenir en compte que algunes particions són molt petites respecte de les altres (un efecte clar és el corpus del Sexenni Liberal, 1869-1874). Així, si l'índex de la IM és alt, els termes comparteixen el context més freqüentment. En el nostre exemple, la IM de tots els casos és molt baixa, per la poca freqüència dels termes en el mateix context respecte de tot el corpus i per l'aparició dels termes de manera independent.

$Tf*idf$ (*term frequency * inverse document frequency*) (Manning i Schutze, 1999) mesura la rellevància d'una paraula en un document respecte d'una col·lecció de documents. En concret, posa en relació la freqüència del terme en un text amb la freqüència de la mateixa paraula en altres textos.

$Tf*idf$ és el resultat de multiplicar dues mesures. En primer lloc, la freqüència del terme (tf) (b) i, en segon lloc, el logaritme de la divisió del nombre total de documents pel nombre de documents que contenen el terme ($idf*tD$) (c).

$$(b) \quad \text{tf}(t, d) = 0.5 + \frac{0.5 \times f(t, d)}{\max\{f(w, d) : w \in d\}}$$

En aquest cas, f indica la freqüència del terme (t) en un document concret (d). Veiem que hi ha una constant en la fórmula; per tant, sempre es donarà un valor mínim (0,5). Aquesta fórmula posa en relació la freqüència del terme en un document amb la freqüència de la paraula més freqüent del document.

$$(c) \quad \text{idf}(t, D) = \log \frac{|D|}{|\{d \in D : t \in d\}|}$$

$\text{Idf}(t,D)$ (c) indica la rellevància d'un terme en un document en relació amb una col·lecció de documents. Aquest valor augmenta si una paraula apareix en un document (d) i no en altres ($D-d$). Aquesta mesura permet, per exemple, detectar termes de llenguatge d'especialitat si l'apliquem a una col·lecció de documents de diferents dominis temàtics. La mateixa mesura podria detectar termes característics d'una varietat dialectal si la col·lecció de documents fos de diferents varietats dialectals.

Finalment, tf^*idf es calcula multiplicant el resultat d'aquestes dues mesures (c).

$$(c) \quad \text{tf}^*\text{idf}(t,d,D) = \text{tf}(t,d) * \text{idf}(t,D)$$

Hem aplicat aquesta fórmula a les particions basades en el règim polític del CTILC utilitzant tres paraules: *dona*, *anarquisme* i *computadora*. La figura 7 ens en mostra els resultats.

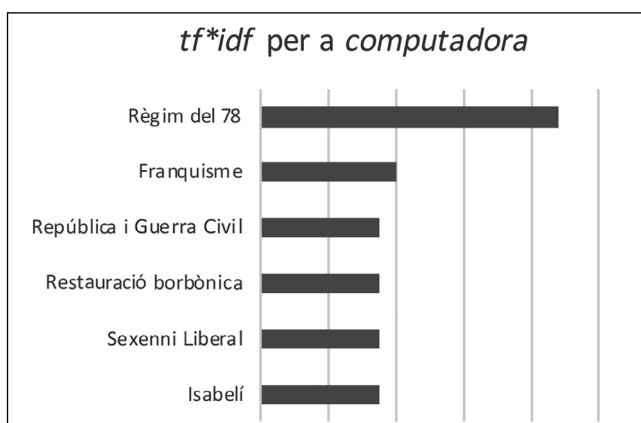
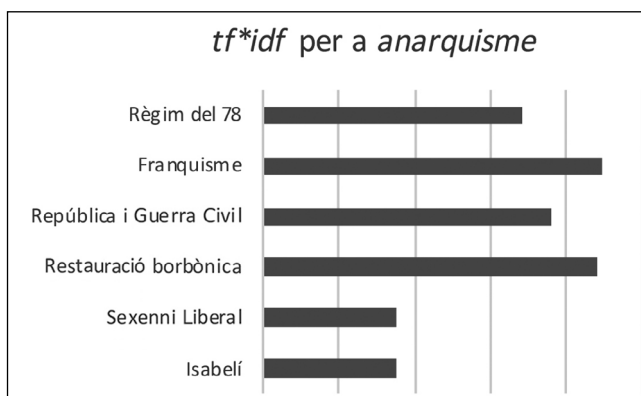
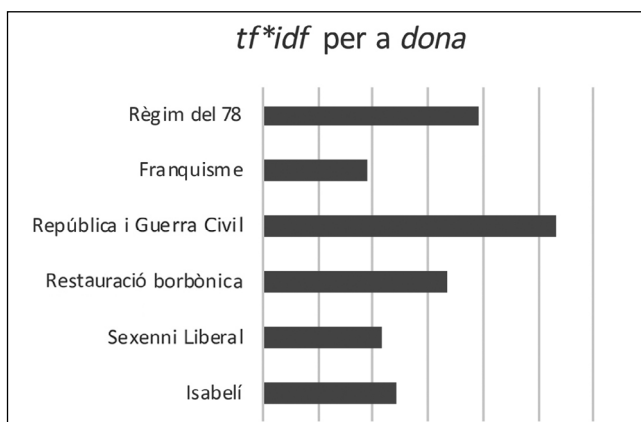


Figura 7. Gràfic dels resultats $tf*idf$ dels termes *dona*, *anarquisme* i *computadora* en les diferents particions del CTILC.

El que indiquen aquests resultats és que les paraules estudiades són més representatives en els períodes en què té més valor la *tf*idf*. Com que les particions són grans i atesa la poca freqüència relativa dels temes, els valors són molts petits, però ja podem captar aquestes diferències mínimes. Observem a *computadora* que la fórmula fa que, encara que no hi hagi ocurrències, sempre es dona un valor (0,5) en les quatre primeres particions, i puja clarament en la darrera partició. Així, els resultats només seran vàlids de manera comparativa dins de la col·lecció de documents, encara que les diferències siguin petites.

3.1. N-grams

Els n-grams són agrupacions seqüencials de les paraules del text; poden ser seqüències de dues paraules (bigrames), de tres (trigrames) o de més. Aquest tractament de les paraules d'un corpus s'ha utilitzat força en lingüística computacional per tal de construir models del llenguatge. Aquests models s'aprenen de manera inductiva des del text i cadascun d'aquests grups es pot associar amb probabilitats d'aparició del grup respecte al volum general del corpus.

Els n-grams permeten extreure associacions de paraules d'un text sense saber d'entrada el que cerquem, només fixant-nos en la freqüència d'aparició d'aquell grup. Ara bé, observar amb quines paraules s'associa una altra paraula pot ajudar-nos a caracteritzar un text i establir diferències amb altres textos. Seguint amb el nostre exemple, podem observar el que ens aporten els n-grams aplicant aquesta tècnica a cadascuna de les particions.

Per exemple, observem els n-grams del verb *minar* en les diferents particions. Segons el *Diccionari de la llengua catalana* de l'Institut d'Estudis Catalans (DIEC2) hi ha dos sentits principals: el de posar un explosiu (1) i el segon, que entenem com una extensió del primer i que és afectar alguna situació o condició (2).

(1) 1 1 v. tr. [MI] [AQ] [DE] Fer una mina o mines (en un terreny, sota una fortificació, etc.). Van minar el mur i el feren saltar amb dinamita.

...

(2) 2 tr. [LC] Preparar la ruïna (d'algú o d'alguna cosa). La tisi li minava la salut. La febre li ha minat les forces.

Com podem veure (figura 8), *minar* en la partició més antiga (1830-1868) ocorre amb objectes de tipus físic («torre», «fonament»), que interpretarem com a possibles objectes directes i instruments («pólvora»), de manera que hi podem

associar de manera clara el sentit 1 del DIEC2. Ja en aquest primer període trobem un sentit figurat que s'associa amb la causa (febre) i que es correspon amb el sentit 2. Si ho comparem amb el període 1930-1940, uns cent anys després, observem que ja totes les ocurrences de *minar* són extensions del sentit físic i, per tant, es relacionen amb el segon sentit.

1830-1868	1930-1940
1 minar torre	1 minar existència
1 minar pólvora	1 minar vida
1 minar fonament	1 minar societat
1 minar febre	1 minar socialdemocràcia

Figura 8. Bigrames del verb *minar* en diferents particions (1830-1868 i 1930-1940).

Com hem vist, aquesta tècnica ens permet veure els fenòmens lèxics de coocurrència i aplicar l'anàlisi estadística a grups de paraules.

Prenem l'exemple del bigrama *cop estat* (recordem que no hi ha paraules funcionals en aquestes particions i, per tant, la preposició *de* no és al text) (taula 2). En la segona partició (corresponent al Sexenni Liberal) veiem que el nombre d'aparicions d'aquest bigrama és baixa.

Període	Ocurrences
1830-1868	1
1869-1874	6
1875-1930	34
1931-1939	19
1940-1975	51
1976-2003	44

Taula 2. Frequència absoluta del bigrama *cop estat* en les diferents particions del corpus.

Ara bé, sabem que el corpus de la segona partició és molt reduït quant al nombre de paraules i, així, la freqüència relativa és més alta que en la resta de particions, que presentem en la figura 9.

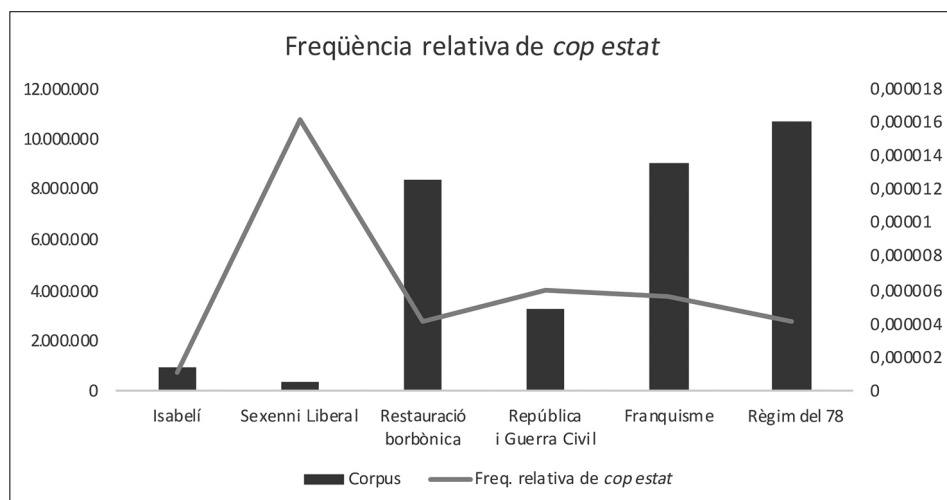


Figura 9. Freqüència relativa de *cop estat* en les diferents particions.

La freqüència relativa ens mostra la importància d'aquest bigrama en aquesta partició, i ho veiem amb relació a la resta, on aquest grup no té tanta presència.

Si sobre el mateix exemple calculem la IM entre *cop* i *estat*, veurem que també aquesta mesura ens indica la importància d'aquest grup en la partició i també en relació amb la resta (figura 10).

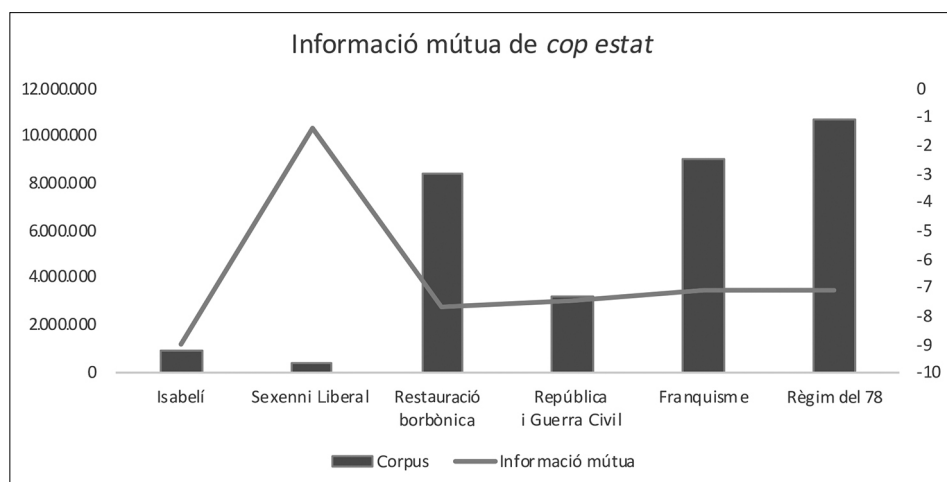


Figura 10. En barres el volum de paraules del corpus i en línia la IM de *cop* i *estat* en les diferents particions basades en règims polítics.

Per acabar, observem el resultat d'aplicar la mesura $tf*idf$ al bigrama *cop estat* en el conjunt de particions. Aquesta mesura compara la freqüència d'aquest bigrama amb la freqüència en la resta de documents. En la figura 11 veiem que, respecte a la resta de documents, la presència d'aquest bigrama és molt més representativa en la segona partició que en les altres.

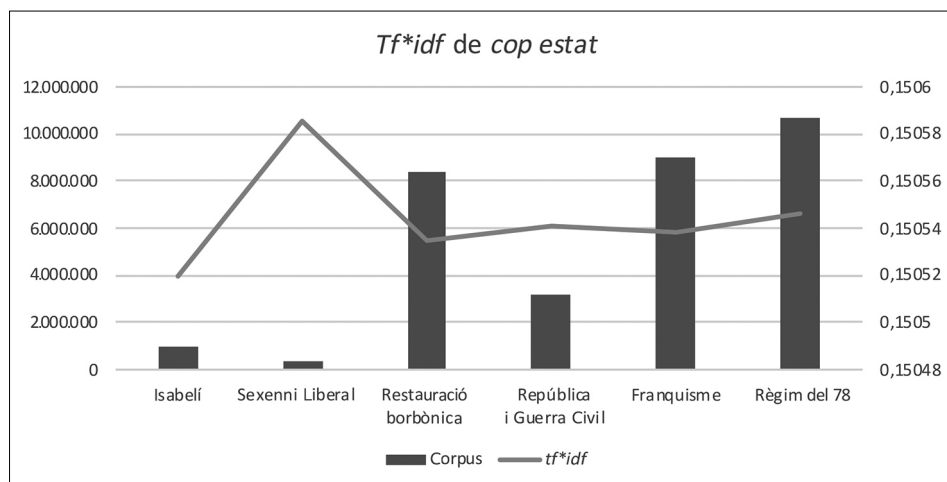


Figura 11. $Tf*idf$ de *cop i estat* en les diferents parts del corpus (partició 1).

Com podem veure, totes tres mesures ens informen de la importància del grup en aquest període respecte dels altres, tant pel nombre d'ocurrències (freqüència), per les freqüències conjunta i independents (IM), com per la seva relació amb la resta de documents ($tf*idf$).

4. APRENTATGE AUTOMÀTIC

El darrer punt que tractarem és l'anàlisi amb tècniques d'aprenentatge automàtic, la línia de recerca dels darrers anys que més bons resultats proporciona per a diferents nivells d'anàlisi lingüística. La base d'aquests mètodes és la inducció de models de llenguatge a partir de corpus de dades lingüístiques de gran volum. En aquest punt tractarem, des d'una visió general, els diferents mètodes utilitzats i l'ús de models basats en xarxes neurals que es coneixen com a aprenentatge profund (*deep learning*).

En un sistema d'aprenentatge supervisat, la inducció o la construcció del model es fa a partir de dades o exemples anotats prèviament en el nivell requere-

rit per la tasca. Això implica que hi ha una dependència d'aquests mètodes respecte del desenvolupament i l'anotació de corpus, ja que per a mètodes supervisats serà necessari un gran volum de dades anotades. Hi ha molts exemples de tasques lingüístiques realitzades amb aquests mètodes (Kotsiantis, 2007).

Hi ha, però, mètodes híbrids entre la supervisió i la no supervisió, sistemes semisupervisats, com, per exemple, les tècniques de *bootstrapping*, és a dir, un aprenentatge aplicat de manera incremental, que utilitza pocs recursos a l'inici i incorpora el coneixement dels exemples que va analitzant (Wang *et al.*, 2007).

Com hem dit, l'aprenentatge no supervisat implica disposar de grans volums de text; ara bé, no cal que aquests textos estiguin anotats, per la qual cosa són mètodes que no requereixen un gran esforç. Una de les tècniques més utilitzades és la classificació automàtica. Els sistemes de classificació són capaços de mesurar les similituds dels exemples a partir d'una sèrie d'atributs —com ara dades contextuals—, representats en una matriu vectorial. Aquestes classificacions es fan utilitzant una mesura de similitud que permet calcular la distància entre els diferents exemples. També fan servir algorismes que proposen diferents procediments per prioritzar la manera de classificar (arbori, *k-means*, etc.) (Manning i Shultze, 1999). N'és un bon exemple l'eina CLUTO (Karypis, 2003).

El resultat no acostuma a ser una classificació perfecta, però sí que es poden veure diferents grups de comportament. En aquest sentit, aquests classificadors s'han fet servir per a estudis lingüístics com ara el del diferent grau de funcionalitat de les preposicions (Reina, 2020) o el de la inducció de models verbals (Gil, 2019).

4.1. Xarxes neurals

L'aprenentatge profund es fa amb un conjunt de mètodes/tècniques que es basen en la capacitat de representar les paraules contextualitzades com a vectors continus en un espai de dimensionalitat reduïda. Aquest mètode és conegut com a *word embeddings*. Un dels sistemes més utilitzats i un dels primers a compartir-se amb la comunitat investigadora és word2Vec (Mikolov *et al.*, 2013). Aquest sistema proporciona una representació única de cada paraula, elaborada a partir dels contextos on ocorre.

VecMap (Artetxe *et al.*, 2017) és un altre exemple d'aplicació d'aquestes tècniques basades en vectors. VecMap permet construir diccionaris bilingües a partir de corpus monolingües (no paral·lels) de dues llengües. Aquest sistema aconsegueix relacionar les dues llengües utilitzant la cartografia dels *embeddings* monolingües.

Aquests sistemes han proporcionat la possibilitat d'aprendre de gran volums de textos sense anotar. Alguns d'aquests sistemes són els basats en arquitectures de *transformers*, entre els quals podem mencionar BERT (Devlin *et al.*, 2018). En BERT trobem models de llenguatge preentrenats que mostren un cert coneixement del món. Els resultats en diverses tasques de comprensió del llenguatge han estat bons, però encara són lluny dels de l'humà (Petroni *et al.*, 2019).

En definitiva, les xarxes neurals són una bona tècnica per a la modelització de sistemes complexos. Es tracta de sistemes que poden guardar informació de manera distribuïda i poden crear la seva pròpia autoorganització de la informació. Són robustos i els models resultants operen en temps real. Permeten resoldre problemes per a tasques de classificació i el reconeixement de patrons i de comportaments. Ara bé, no tenen capacitat per interpretar els seus resultats i per a segons quines tasques es necessita molt volum de dades. S'han aplicat a múltiples camps, com ara la física, la computació, les neurociències, la salut i la lingüística, entre d'altres (Rivas-Asanza *et al.*, 2018). En el terreny de la lingüística, aquests sistemes aplicats a les paraules donen suport a la hipòtesi distribucionalista de Harris (1954) i donen lloc a la hipòtesi neodistribucionalista; és a dir, de manera simple, les paraules similars semànticament tenen una distribució similar.

Per mostrar un exemple de com treballen aquests mètodes i de la seva utilitat en l'estudi de la variació lèxica, utilitzarem el word2vec amb les particions presentades a l'inici de la presentació. Donada una paraula, el word2vec ens proporciona la paraula més similar contextualment dels textos entrenats. És a dir, per tal de poder treballar hem de tenir inicialment un model ja entrenat o bé hem d'entrenar-ne un. En el nostre cas, hem construït un model per a cada una de les particions presentades. Per exemplificar la funció del word2Vec en aquesta recerca vegem l'exemple de la figura 12, on presentem les paraules més similars a *rei* i *obrer* segons aquest sistema.

Word: rei Position in vocabulary: 148		Word: obrer Position in vocabulary: 211	
Word	Cosine distance	Word	Cosine distance
monarca	0.640851	treballador	0.632402
príncep	0.589831	patronal	0.555207
conqueridor	0.574684	afiliat	0.512385
ambaixador	0.559783	assalariat	0.492871

Figura 12. Paraules més similars a *rei* (esquerra) i *obrer* (dreta); a la dreta de la paraula s'expressa un índex de la distància semàntica entre el terme cercat i el terme resultant.

El word2vec proporciona una mesura de similitud basada en el cosinus, amb la qual podem extreure les paraules més similars per a cada període sempre a partir d'una paraula de cerca. En general proporciona paraules relacionades semànticament i paradigmàticament amb els termes de cerca, però també pot proporcionar algunes relacions sintagmàtiques.

Aquesta manera d'analitzar les paraules del corpus ens permet veure diferències en el sentit de les paraules, és a dir, podem detectar canvis de camps semàntics o de dominis, la qual cosa podria indicar un canvi en la prototipicitat del sentit del terme o bé un desplaçament semàntic. Com a exemple, observem les paraules similars a *crema* en diferents períodes temporals en la figura 13.

part 6	1940-1959	part 7	1960-1979	part 8	1980-1999	part 9	2000-2003
crema vainilla 0.627693620177		crema massapà 0.550693126224		crema loció 0.677436115464		crema flam 0.540937384436	
crema praliné 0.57327975206		crema confitura 0.527812218339		crema pomada 0.629036495208		crema massapà 0.539473610399	
crema marrasquí 0.542650350861		crema panellet 0.526212605698		crema lanolina 0.627706701876		crema tortell 0.518544241054	
crema natilla 0.537019046833		crema bescuit 0.521716122248		crema xampú 0.610563599087		crema llard 0.513452963133	
crema glacé 0.535809438549		crema nata 0.504369935241		crema liniment 0.610199470011		crema torró 0.505032204	
crema mermelada 0.533193578165		crema mató 0.499874275548		crema suavitzant 0.584628342399		crema pistatxo 0.493752448042	
crema bescuit 0.515020986491		crema melindro 0.486986051353		crema hidratant 0.583527739421		crema iogurt 0.490196818518	

Figura 13. Les set primeres paraules similars a *crema* en diferents períodes. Observem que en general domina el sentit gastronòmic, però en el període 1980-1999 s'imposa el domini de la higiene com el sentit més utilitzat en el corpus.

Com dèiem, aquesta tècnica també ens permet localitzar nous sentits de paraules ja existents. Observem la figura 14, on es llisten les paraules similars a *premsa* en diferents períodes. Podem observar que hi ha un nou sentit en la partició corresponent a 1870-1889, sentit que ja serà el dominant en les particions següents.

Part 1	1830- 1849	Part2	1850-1869	Part3	1870-1889	Part4	1890-1919
premsa <u>cloruro</u> 0.8373		premsa cartró 0.7336		premsa redactor 0.6182		premsa periòdic 0.5918	
premsa transparent 0.8301		premsa estanc 0.7253		premsa periòdic 0.6099		premsa propaganda 0.5367	
premsa guix 0.8153		premsa davantal 0.7240		premsa redacció 0.6090		premsa redactor 0.5344	
premsa descompondre 0.8072		premsa sucar 0.7005		premsa satíric 0.5442		premsa setmanari 0.5326	
premsa sòlid 0.7965		premsa emplenar 0.7004		premsa secretaria 0.5433		premsa quinzenari 0.5212	
premsa <u>franel·la</u> 0.7893		premsa enganxar 0.6984		premsa ateneu 0.5221		premsa míting 0.4980	
premsa copal 0.7782		premsa puro 0.6954		premsa censurar 0.5208		premsa periodisme 0.47628	

Figura 14. Les set primeres paraules similars a *premsa* en diferents períodes.

Un altre exemple de com aquestes eines ens ajuden a explorar els canvis lèxics el trobem en la paraula *hospital*, que a l'inici del segle xx (període 1890-1919) es fa servir en contextos més propers a la idea de beneficència, mentre que a mitjan segle xx (període 1940-1959) se situa en uns contextos que són propers a la idea d'hospital sanitari (vegeu la figura 15).

1890-1919	1940-1959
hospital misericòrdia 0.4992	hospital dispensari 0.657
hospital convent 0.4638	hospital anticancerós 0.6005
hospital orde 0.4452	hospital hospitalitzar 0.5953
hospital <u>quartel</u> 0.4436	hospital sanatori 0.5860
hospital monestir 0.4413	hospital ambulància 0.5767
hospital hospitaler 0.4356	hospital clínica 0.5671
hospital dignatari 0.4351	hospital cancerós 0.5669
hospital beneficència 0.4335	hospital hospitalització 0.5229
hospital hospici 0.4299	hospital infermeria 0.5169
hospital rectoral 0.4296	hospital asil 0.5163

Figura 15. Paraules similars a *hospital* en dos períodes diferents del segle xx.

En resum, aquestes tècniques ens permeten explorar la semàntica de les paraules a partir de la similitud o no del context on ocorren. Aquest estudi es pot fer a partir de corpus no anotats o bé de corpus anotats, la qual cosa proporcionarà resultats potser més acurats, ja que anotar el corpus permet filtrar almenys els termes homònims. Es tracta d'una metodologia que ens pot ajudar a estudiar els comportaments semàntics i que, aplicada sobre corpus de diferents èpoques, ens permet analitzar-ne la variació.

5. CONCLUSIONS

Les diferents mesures o tècniques presentades per al tractament massiu de textos aplicats a corpus diacrònics semblen funcionar positivament. Volem destacar la necessitat de disposar de corpus de la llengua de gran volum per tal de poder aplicar mesures estadístiques i mètodes d'aprenentatge automàtic.

També hem intentat mostrar els avantatges de l'ús d'aquestes tècniques per poder tractar el nivell semàntic de les llengües. Així, aquestes eines ens permeten descobrir canvis en la freqüència de grups de paraules, mesurar la rellevància d'aquests grups a través de la comparació amb altres corpus (particions)

i veure canvis contextuais en el temps i desplaçaments semàntics. Creiem que les aportacions d'aquesta metodologia són la possibilitat de l'exploració semàntica dels corpus i van una mica més enllà de la forma de la paraula.

Treballar amb aquests mètodes per a la detecció de polisèmies, basades en el context de la paraula i en el seu domini, és un dels reptes futurs d'aquesta recerca. També trobem interessant la construcció d'un sistema d'anàlisi de corpus que integri totes aquestes funcions per tal de poder descobrir, a partir de les mesures i els resultats, fenòmens de canvi.

Com a projecte de futur, volem aplicar aquestes mesures a tot el corpus per tal de no haver de cercar els termes que s'han destudiar, tot confiant que aquestes eines ens proporcionin el vocabulari interessant per a l'estudi de la variació.

AGRAÏMENTS

Volem agrair la cessió del corpus per fer aquest estudi a Joaquim Rafel i a Joan Soler per al CTILC, a Joan Torruella per al CICA i a Joan Veny i Àngels Massip per al corpus Scripta.

REFERÈNCIES BIBLIOGRÀFIQUES

- ARTETXE, Mikel; LABAKA, Gorka; AGIRRE, Eneko (2017). «Learning bilingual word embeddings with (almost) no bilingual data». A: *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, vol. 1, pàg. 451-462.
- DEVLIN, Jacob; CHANG, Ming-Wwei; LEE, Kenton; TOUTANOVA, Kristina (2018). *BERT: Pre-training of deep bidirectional transformers for language understanding*. arXiv preprint arXiv:1810.04805. DOI: 10.48550/arXiv.1810.04805.
- GIL VALLEJO, Lara (2020). *Exploiting verb similarity for event modelling*. Tesi doctoral. Barcelona: Universitat Oberta de Catalunya, Internet Interdisciplinary Institut (IN3). <http://hdl.handle.net/10803/668907>.
- HARRIS, Zellig (1954). «Distributional Structure». *Word*, vols. 2-3, núm. 10, pàg. 146-162.
- KARYPIS, George (2003). *CLUTO: A clustering toolkit*. Universidad de Minnesota, Technical Report, 02-017.A <http://glaros.dtc.umn.edu/gkhome/cluto/cluto/overview>.
- KOTSIANTIS, S. B. (2007). «Supervised machine learning: a review of classification techniques». A: MAGLOGIANNIS, I. et al. (eds.). *Emerging artificial intelligence applications in computer engineering*. Amsterdam: IOS Press, pàg. 3-24.
- MANNING, Chris; SCHÜTZE, Hinrich (1999). *Foundations of statistical natural language processing*. Cambridge, MA: MIT Press.

- MIKOLOV, Tomas; CHEN, Kai; CORRADO, Greg; DEAN, Jeffrey (2013). «Efficient estimation of word representations in vector space». A: *Proceedings of ICLR Workshops Track*. <https://arxiv.org/abs/1301.3781>.
- PETRONI, Fabio; ROCKTÄSCHEL, Tim; RIEDEL, Sebastian; LEWIS, Patrick; BAKHTIN, Anton; WU, Yuxiang; MILLER, Alexander. (2019). «Language models as knowledge bases?». A: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Hong Kong: Association for Computational Linguistics, pàg. 2463-2473.
- REINA, Francesc (2020). *Significado, distribución y frecuencia de la categoría preposicional en español. Una aproximación computacional*. Tesis doctoral. Barcelona: Universitat de Barcelona.
- RIVAS-ASANZA, Wilmer; MAZON-OLIVO, Bertha; MEJÍA, Fernando (coords.) (2018). «Generalidades de las redes neuronales artificiales». *Redes neuronales artificiales aplicadas al reconocimiento de patrones*. www.researchgate.net/publication/327703478_Capitulo_1_Generalidades_de_las_redes_neuronales_artificiales [consulta: 21-7-2020].
- WANG, Wen; HUANG, Zhongqiang; HARPER, Mary (2007). «Semi-supervised learning for part-of-speech tagging of Mandarin transcribed speech». *Acoustics, Speech, and Signal Processing – ICASSP '07*, Honolulu, 2007, pàg. 137-140. DOI: 10.1109/ICASSP.2007.367182.

Corpus

- Corpus Textual Informatitzat de la Llengua Catalan-CTILC*. Joaquim Rafel Fontanals, director, i Joan Soler, coordinador. <https://ctilc.iec.cat/scripts/IniCredits.asp>.
- Corpus Informatitzat del Català Antic-CICA* (2009). Joan Torruella, director, Manel Pérez Saldanya i Josep Martines. <http://lexicon.uab.cat/cica>.
- Corpus SCRIPTA – Dialectologia Catalana* (2009). Joan Veny i Àngels Massip. www.scripta-dialectologia.ub.edu/scripta-online.

Diccionaris

- Diccionari de la llengua catalana* (DIEC2). Barcelona: Institut d'Estudis Catalans. <https://dlc.iec.cat>.

Cap a una quantificació de l'evolució semàntica del lèxic mitjançant tècniques basades en corpus diacrònics i word2vec

Gemma Bel-Enguix

Universitat de Barcelona

Universidad Nacional Autónoma de México*

1. INTRODUCCIÓ

Les tècniques desenvolupades per la lingüística en els darrers cent anys han incidit de manera específica en els mètodes anomenats «sincrònics», que estudien estadis determinats d'una llengua com a entitat estàtica. Aquest plantejament ahistoricista està sent qüestionat en els darrers anys (Cilliers, 2013; Bastardas-Boada, 2016), ja que la ciència actual aproxima els fenòmens de manera cada cop més unificada. D'altra banda, des de la teoria de la complexitat, nou paraigua científic que proporciona les eines necessàries per dur a terme aquesta visió unificada, l'evolució és un dels components de tot sistema dinàmic adaptatiu (Holland, 1992; Gell-Mann, 1995). En la complexitat no es poden separar els elements constitutius d'un tot: existeix un teixit interdependent, interactiu i retroactiu entre les parts i el tot, el tot i les parts.

Com diu Massip-Bonet (2018), la perspectiva temporal en els estudis lingüístics aporta, tant a la recerca com a la lingüística aplicada, el factor d'ubicació, d'emmarcament dins la realitat. Dona compte de l'escenari geogràfic, socioeconòmic i ecològic en què una llengua s'ha format. Quan es parla d'evolució en biologia, les contingències històriques i ambientals es consideren determinants. El mateix ocorre en lingüística: una llengua no deixa de ser un sistema com altres (oberts i dinàmics) que són estudiats per les ciències.

Qualsevol acció és un canvi d'estat, i un estat és un conjunt d'accions potencials. Quan qualsevol d'aquestes accions s'executa, l'estat canvia. Per això la llengua està en evolució constant, perquè s'executa constantment.

La variació sistèmica és el canvi (en curs) causat per accions (en el cas que ens ocupa, lingüístiques). Es tracta d'un procés, en certa manera, cec, ja que cap

* Agraïxo el suport del Ministeri de Ciència i Innovació mitjançant el programa María Zambrano, finançat per la Unió Europea, Next Generation EU.

agent, encara que tingui un objectiu, no pot preveure totes les conseqüències de les seves actuacions.

La variació lingüística resulta de captar una imatge estàtica de la llengua projectada en el contínuum espai-temps. Per la seva pròpia naturalesa, doncs, només pot ser estudiada des d'un punt de vista integratiu i tenint en compte la dimensió temporal.

La sincronia, llavors, és un pla dins la diacronia. I la diacronia només és una perspectiva vàlida quan es tenen dades de diferents talls sincrònics.

Aquest article parteix de la idea que els corrents actuals d'estudi del llenguatge natural, que es basen en la compilació de grans corpus de text i l'anàlisi mitjançant mètodes estadístics, poden servir per interpretar des d'un altre punt de vista el tema de l'evolució del llenguatge.

Per dur a terme el treball s'empraran tècniques senzilles d'anàlisi de freqüències i d'altres basades en word2vec, un sistema de vectors entrenat amb xarxes de neurones.

Es proposa, doncs, un estudi sincrònic basat en corpus de diferents períodes dels segles XIX i XX per fer-ne després una interpretació diacrònica. Per això, en la secció 2 es fa un breu repàs del marc general d'estudi de la semàntica i el canvi semàntic dins del processament del llenguatge natural. La secció 3 se centra en l'explicació general de la metodologia emprada en la preparació dels corpus i la generació dels vectors. En la secció 4 s'expliquen els experiments basats en freqüències, mentre que la secció 5 desenvolupa els experiments basats en vectors. Finalment, la secció 6 presenta les conclusions i direccions futures del treball.

2. ESTAT DE L'ART

Mentre que algunes branques de la lingüística han emprat des de fa dècades la formalització matemàtica i la informàtica per desenvolupar eines efectives d'anàlisi, la semàntica va trigar a trobar mètodes eficients per representar el significat de manera computacional. Una de les raons d'aquest endarreriment ha estat la necessitat d'una gran capacitat de processament per dur a terme la implementació d'idees que havien sorgit a mitjan segle XX i que requerien tasques de comparativa textual.

Les tècniques per capturar el significat de lèxic, oracions i documents es basen en la intuïció de Harris (1954) sobre la teoria distribucional. Aquesta idea estableix, principalment, que dues paraules que funcionen en els mateixos contextos són similars (Firth, 1957). A partir d'aquí es va iniciar el treball amb *word embeddings* (Clark, 2015) i més tard, amb l'esclat i l'expansió dels mèto-

des basats en xarxes neurals, van sorgir els vectors entrenats amb aquests models, com word2vec (Mikolov, 2013).

El bon funcionament demostrat per word2vec ha impulsat la creació d'altres vectors que s'han mostrat més eficients en alguns contextos, com GloVe (Pennington *et al.*, 2014) o FastText (Bojanowski *et al.*, 2016). En els darrers dos anys han aparegut els *word embeddings* contextuals —ELMO (Peters *et al.*, 2018)— i els models basats en *transformers*, que inclouen mecanismes d'atenció. BERT (Devlin *et al.*, 2018), entrenat amb enormes quantitats de text, és l'exemple més conegut de model de llenguatge basat en la tecnologia *transformer* (Vaswani *et al.*, 2017), que ha millorat els resultats obtinguts pels originals word2vec en moltes tasques de processament del llenguatge natural.

Mitjançant els vectors es poden aproximar fenòmens como la sinonímia, la polisèmia o la dispersió semàntica. Però aquestes eines no poden, per si mateixes, codificar de cap manera l'evolució del llenguatge.

Per tal d'abordar els canvis semàntics en el temps, o l'evolució semàntica de mots determinats, s'ha proposat de fer corpus d'èpoques diferents de la llengua i estudiar-los amb les mateixes tècniques que s'apliquen als recursos sincrònics, per després fer-ne una comparació.

K. Gulordava i M. Baroni (2011) van utilitzar l'eina n-gram de Google per resseguir l'evolució d'un mot. Un pas més van fer Kulkarni *et al.* (2015) en entrenar vectors en corpus de diferents èpoques per després estudiar-ne els resultats de manera conjunta. Hamilton *et al.* (2016) treballen els vectors en corpus diacrònics per descobrir les lleis del canvi semàntic.

Aquest article és una aplicació inicial d'algunes d'aquestes tècniques a un corpus diacrònic del català, que hem dividit en sis etapes diferents. D'aquesta manera es vol comprovar la idoneïtat de la metodologia en recursos i llengües diferents de l'anglès.

3. METODOLOGIA

Aquest estudi s'ha dut a terme sobre el Corpus Textual Informatitzat de la Llengua Catalana (CTILC),¹ dirigit per Joaquim Rafel, que conté textos en català des de 1833 a 2008 i que a finals de 2020 inclou setanta-dos milions de mots.

Les dades amb què treballem inclouen sis períodes històrics que van des del 1843 fins al segle XXI. Hem denominat així els períodes: isabelí (1843-1868),

1 <https://ctilc.iec.cat>.

Sexenni Liberal (1868-1874), Restauració borbònica (1875-1930), República i Guerra Civil (1931-1939), franquisme (1940-1978) i Règim del 78 (1978-). Per als detalls sobre el tractament del corpus i l'entrenament dels models de llenguatge, així com molts dels resultats lèxics que se n'han extret, vegeu els apartats 3 i 4 del capítol de Castellón en aquest llibre.

Aquestes etapes han estat establertes per anar comprovant com l'evolució del llenguatge va de bracet de qüestions polítiques i socials, sobretot pel que fa al lèxic. Hi ha, també, criteris de corpus. Per exemple, encara que els anys de la Segona República (1931-1936) són molt diferents dels de la Guerra Civil, no hi havia un volum prou gran de paraules en aquests tres anys, en el moment de la realització dels experiments, per poder entrenar models vectorials. La taula 1 (vegeu l'apartat 2 la taula 1 del capítol de Castellón en aquest llibre) mostra la distribució de paraules per a cada un dels sis períodes en què s'ha dividit el treball.

Període	Anys	Paraules
Isabelí	1843-1868	836.857
Sexenni Liberal	1868-1874	334.155
Restauració borbònica	1875-1930	6.470.870
República i Guerra Civil	1931-1939	2.888.576
Franquisme	1949-1978	8.186.055
Règim del 78	1978-1988	5.378.649

Taula 1. Configuració del corpus.

El treball que presentem es basa en un estudi de les freqüències lèxiques per a cada un d'aquests períodes (secció 4) i en una anàlisi de similitud semàntica (secció 5).

Sobre cada una d'aquestes particions del corpus s'ha entrenat un sistema de vectors word2vec (Mikolov, 2014), que ha servit per establir una comparativa que dona una idea de l'evolució dels mots.

4. EXPERIMENTS BASATS EN FREQUÈNCIES

En primer lloc es presenta un estudi general de freqüències per a cada una de les etapes que s'han establert. Per dur-lo a terme s'han eliminat les paraules funcionals i també alguns verbs que poden exercir d'auxiliars, com *ésser* o *haver*. El primer pas ha estat fer una anàlisi del corpus mitjançant FreeLing (Carreras

et al., 2004) per destriar-ne tres grups, sobre els quals s'ha enfocat el treball: substantius (secció 4.1), adjectius (secció 4.2) i adverbis (secció 4.3). Després, per a cada període històric s'han agafat les cinquanta instàncies més freqüents en cadascuna d'aquestes categories.

4.1. Substantius

Pel que fa als substantius, la taula 2 mostra el resultats comparatius dels quinze primers elements més freqüents. No es dona el nombre absolut d'ocurrències perquè no és significatiu a causa del desequilibri existent entre els conjunts de dades. Si es fa un seguiment del substantiu més utilitzat en cada període, es troba *Déu* durant els anys del regnat d'Isabel II, mentre que durant el Sexenni Liberal la paraula amb més ocurrències en el corpus és *senyor*, i durant la Restauració borbònica el primer lloc l'obté *home*, igual que durant el franquisme. Durant la Guerra Civil el genèric *cosa* guanya a *home*. Finalment, durant el Règim del 78 la paraula més utilitzada és *any*.

Període/ rànkung	Isabelí	Sexenni Liberal	Restauració borbònica	República – Guerra Civil	Franquisme	Règim del 78
1	Déu	Senyor	Home	Cosa	Home	Any
2	Dia	Dia	Cosa	Home	Any	Dia
3	Senyor	Terra	Dia	Dia	Cosa	Home
4	Terra	Déu	Any	Any	Dia	Cosa
5	Cosa	Cosa	Temps	Casa	Vida	Part
6	Temps	Home	Terra	Vida	Temps	Manera
7	Fill	Pare	Vida	Senyor	Vegada	Temps
8	Pare	Casa	Senyor	Vegada	Manera	Vida
9	Rei	Temps	Part	Temps	Déu	Casa
10	Home	Cor	Casa	Part	Casa	Vegada
11	Any	Fill	Déu	Manera	Fet	Cas
12	Part	Any	Deure	Poble	Part	Terra
13	Casa	Vegada	Poble	Terra	Terra	Lloc
14	Cor	Paraula	Vegada	Mà	Món	Senyor
15	Don	Vida	Manera	Lloc	Poble	Forma

Taula 2. Rànkung dels quinze substantius més usats en cada un dels períodes estudiats.

Aquests resultats s'han de prendre com a exemples il·lustratius i de cap manera són conclusius. Però sí que es poden observar alguns fets, que fem notar a continuació.

El substantiu *Déu*, el més emprat en el període isabelí, apareix entre els quinze primers en quatre dels sis períodes, excepte en la Guerra Civil i el Règim del 78. *Senyor*, en el primer lloc en el Sexenni Liberal, és un terme ambigu, ja que pot tenir connotacions polítiques o religioses. Apareix en totes les èpoques excepte durant el franquisme. *Home*, en canvi, el substantiu més freqüent durant l'època borbònica i el franquisme, és present a totes les columnes. El terme agafa, en algunes ocasions, el significat d'humà i de vegades s'interpreta com a humà masculí. És notable, en canvi, que *dona* no apareix ni un sol cop en aquesta taula. De manera anàloga i tenint en compte la suposada importància social de la figura de la mare, aquest mot no apareix ni una vegada en la llista de les quinze primeres paraules en cap època, mentre que sí que ho fan *pare* (en el període isabelí i el Sexenni Liberal) i *fill* (en el Sexenni Liberal).

Finalment, *any*, la paraula més comuna durant el Règim del 78, es troba també a les sis columnes. Això es pot considerar normal perquè el mot es fa servir de manera recurrent en les datacions de documents i en fonts periodístiques, per exemple. Però és notori que el segon substantiu més emprat en el Règim del 78 sigui *dia*. Sembla que la preocupació temporal és recurrent en aquest moment.

Com a curiositats, el mot *rei* apareix només durant el període isabelí, mentre que *cor* s'aprecia durant el Sexenni Liberal i *món* només està present en els primers llocs del rànquing durant el franquisme. Una altra paraula que emergeix durant un temps és *poble*, que es manté durant la Restauració borbònica, la República i la Guerra Civil i el franquisme en la llista de les més freqüents, fet que reflecteix l'aparició de noves idees.

A continuació es fa una comparació entre les sis etapes. S'han agafat els cinquanta substantius més freqüents de cada etapa i s'han filtrat els que són comuns als sis períodes. La llista que s'ha obtingut és la següent: *nit, hora, aigua, lloc, part, ull, dona, mare, mà, punt, dia, pare, vida, fill, cas, senyor, casa, paraula, manera, déu, vegada, terra, cosa, temps, poble, any, gent, nom, món, home*. Aquesta llista no està ordenada, ja que el rànquing que ocupen aquests mots és diferent en cada una de les etapes temporals que tenim en compte. Com que el nombre de *tokens* de cada un dels corpus és divers, per tal d'afavorir la comparativa entre fonts desequilibrades s'ha dut a terme una normalització, consistent a dividir el valor d'ocurrència de cada paraula pel nombre d'aparicions de la paraula més freqüent. D'aquesta manera, en cada època hi ha una paraula (almenys) amb valor 1, mentre que el valor de les altres oscil·la entre 0 i 1. El resultat es mostra en la figura 1.

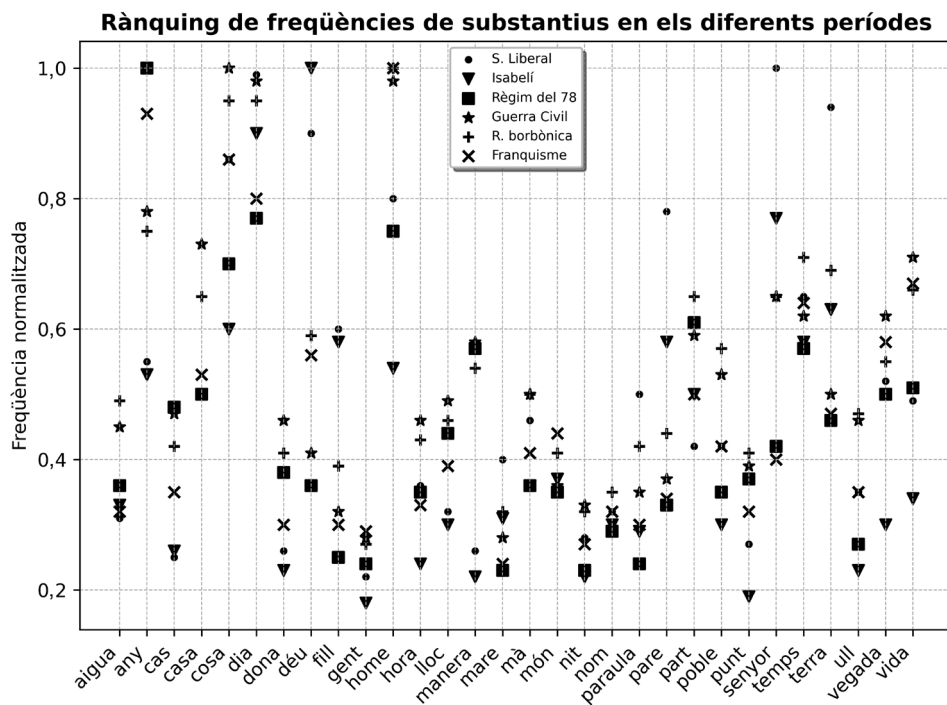


Figura 1. Rànking normalitzat de freqüències dels substantius comuns entre els cinquanta més emprats en els diferents períodes.

4.2. Adjectius

Pel que fa als adjectius, s'ha seguit la mateixa metodologia que en el cas dels substantius. Els resultats dels quinze adjectius més comuns es mostren en la taula 3. La variació entre aquesta classe de paraules és més petita que la que s'ha observat en els substantius. Els primers llocs del rànkning estan ocupats per *bo*, *mateix* i *sant*.

En una llista d'adjectius bàsics amb moltes repeticions crida l'atenció l'aparició de mots únics, com *plural* en el Sexenni Liberal i *social* en el Règim del 78.

Un fet interessant és l'aparició de *català* per primer cop durant el període de la República i la Guerra Civil. El mot se situa, durant el franquisme i el Règim del 78, entre els cinc primers.

Període/ rànquing	Isabelí	Sexenni Liberal	Restauració borbònica	República – Guerra Civil	Franquisme	Règim del 78
1	Bo	Bo	Mateix	Gran	Mateix	Mateix
2	Mateix	Mateix	Gran	Mateix	Gran	Gran
3	Sant	Sant	Bo	Primer	Bo	Bo
4	Gran	Gran	Sant	Bo	Nou	Nou
5	Mal	Pobre	Sol	Sol	Català	Català
6	Primer	Sol	Fi	Sant	Sant	Sant
7	Sol	Fi	Petit	Nou	Sol	Petit
8	Dret	Mig	Mig	Català	Petit	General
9	Present	Nou	General	Petit	Propi	Cert
10	Nou	Present	Dur	Dret	General	Diferent
11	Mig	Plural	Pobre	Mal	Mal	Llarg
12	Pobre	Millor	Alt	General	Dur	Propi
13	Millor	Jove	Bell	Fi	Vell	Important
14	Major	Últim	Llarg	Millor	Humà	Social
15	Últim	Petit	Propi	Mig	Mig	Segon

Taula 3. Resultats dels quinze adjectius més comuns en cada època.

Com en el cas dels substantius, s'han destriat tots els que es troben entre els cinquanta primers a totes les èpoques. El conjunt final en aquest cas només consta de quinze mots: *vell, mateix, mig, petit, natural, alt, llarg, general, antic, clar, gran, bo, major, sant, fort*. Això indica que en català la diferència entre els adjectius més emprats i la resta és més gran que en el cas dels substantius. És a dir, la dispersió lèxica per als substantius és més alta.

La figura 2 mostra el resultat amb les freqüències normalitzades:

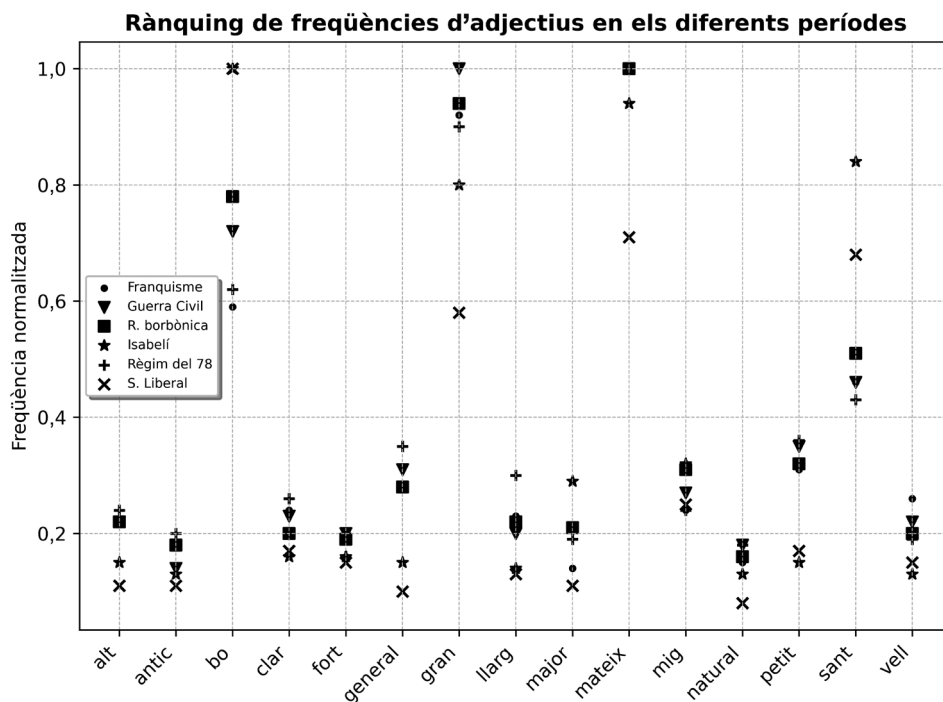


Figura 2. Resultats comparatius dels adjectius que són comuns entre els cinquanta primers en cada una de les èpoques.

4.3. Adverbis

La mateixa metodologia seguida amb els substantius i els adjectius s'ha emprat amb els adverbis. La taula 4 mostra els resultats per a cada etapa en què ha estat dividit el corpus.

Període/ rànquing	Isabelí	Sexenni Liberal	Restauració borbònica	República – Guerra Civil	Franquisme	Règim del 78
1	No	No	No	No	No	No
2	Ja	Ja	Ja	Ja	Ja	Ja
3	Bé	Encara	Encara	També	També	També
4	També	Sempre	Primer	Encara	Encara	Primer
5	Sempre	També	També	Ara	Primer	Encara
6	Ara	Primer	Ara	Sempre	Ara	Ara
7	Encara	Mai	Sempre	Ben	Sempre	Sempre
8	Sí	Sí	Ben	Pas	Només	Només
9	Ben	Ben	Pas	Mai	Pas	Ben
10	Mai	Sols	Bé	Bé	Ben	Sí
11	Poc	Avui	Mai	Sí	Mai	Bé
12	Llavors	Força	Mal	Només	Bé	Pas
13	Avui	Poc	Sí	Avui	Potser	Mai
14	Solament	Potser	Avui	Potser	Sí	Avui
15	Tampoc	Llavors	Millor	Aleshores	Avui	Potser

Taula 4. Resultats dels quinze adverbis més comuns en cada època.

No sembla que hi hagi res rellevant en els resultats de la taula 4. Com mostra la figura 3, la distància entre els dos adverbis més comuns, *no* i *ja*, i la resta és molt gran.

Per elaborar la figura 3 s'han agafat també tots els adverbis comuns entre els cinquanta primers en totes les etapes considerades. El conjunt inclou: *ben*, *ja*, *no*, *encara*, *tard*, *demà*, *només*, *també*, *avui*, *sí*, *aleshores*, *avall*, *potser*, *sovint*, *llavors*, *endavant*, *solament*, *tampoc*, *sempre*, *sobretot*, *amunt*, *mai*, *aviat*, *quasi*. La partícula negativa és a molta distància de *ja*, *encara* i *també*, que també tenen una representació important dins el corpus.

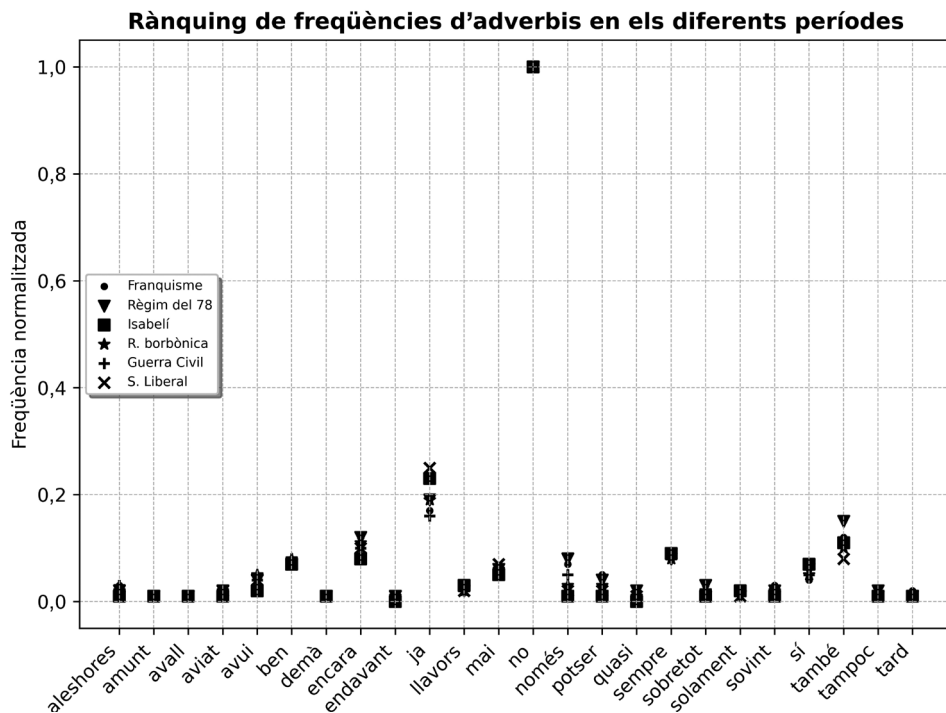


Figura 3. Representació normalitzada de les freqüències dels adverbis en els diferents períodes.

5. EXPERIMENTS BASATS EN VECTORS

Aquestes proves s'han basat en els vectors word2vec entrenats sobre cada un dels subcorpus corresponents a diferents etapes en les quals han estat eliminats els noms propis. Per a aquest experiment no s'ha tingut en compte la secció del Sexenni Liberal ja que, com s'aprecia en la taula 1, el nombre de paraules és substancialment menor que el dels altres períodes, de manera que segurament els vectors tindrien una qualitat molt pitjor.

5.1. Tècniques de visualització

Per tal de representar els vectors en un pla, cal fer-ne abans una reducció de la dimensionalitat. Aquest procés està implementat pel mètode t-SNE (Maaten i Van der Hinton, 2008), que es pot importar des de la llibreria de Python Scikit-learn (Pedregosa *et al.*, 2011). Una característica important i a la vegada bas-

tant imprevisible d'aquest mètode és la perplexitat. Ajustant aquest paràmetre es poden aconseguir representacions ben diferents d'un mateix grup de vectors. Els gràfics que es mostren en aquest article tenen la perplexitat fixada en 40.

En el cas que ens ocupa hi ha un problema fonamental: la dificultat de fer una representació conjunta de vectors que han estat entrenats en diferents corpus per separat. En principi, el sistema word2vec que emergeix de cada conjunt de dades no és comparable. És per això que si es representa una paraula, com ara *dona*, en un pla a partir del vector que sorgeix en l'entrenament en cada un dels subcorpus, les coordenades se situaran en extrems diferents del pla. Però això no està causat per la seva distància semàntica, sinó perquè els espais vectorials de cada conjunt no coincideixen.

Aquesta dificultat va ser abordada per Hamilton *et al.* (2016) i també per Kulkarni *et al.* (2015). Els primers suggereixen dues estratègies per poder tractar les diferents èpoques històriques d'una llengua amb espais euclidians diferents. La primera és evitar els vectors de baixa dimensionalitat, com van fer Gulordava i Baroni (2011). La segona consisteix a implementar tècniques d'alineament. Kulkarni *et al.* (2015) opten per la segona opció, i aquesta també és la que ha estat desenvolupada en el nostre treball. Els autors defineixen el procés d'alineament com a equivalent d'un problema de regressió segmentada, o per trossos. Python té una llibreria, anomenada *pwlfit*,² que implementa aquesta tècnica. Tot i que la solució és costosa computacionalment, sembla que s'adapta a les necessitats del mètode de recerca.

Aquesta metodologia permet agafar un grup de paraules que es consideren relacionades amb *dona* i representar-les en un pla d'una època determinada. Després es fa el mateix amb els vectors de *dona* corresponents als altres conjunts de dades, tot esperant que se situïn propers a les paraules que han estat el seu context en cada moment històric. Però, per a això, aquests vectors que representen el mot *dona* s'han d'alinejar mitjançant la tècnica de la regressió segmentada amb l'espai de les paraules que han estat el seu context.

En tots els resultats que es presenten a partir d'ara s'han dut a terme aquestes transformacions.

2 <https://github.com/google/pwlfit>.

5.2. Desplaçament semàntic de mots

Per exemplificar aquest mètode es pot agafar la paraula *article*, que és part dels cognats *article-artell*, l'evolució històrica del qual pot ser interessant d'estudiar.

El mètode consisteix en els passos següents. Primer s’han consultat amb `word2vec` els deu ítems lèxics més relacionats amb *article* en cada un dels conjunts de vectors. Amb això s’ha construït un conjunt de quaranta-cinc paraules (s’han eliminat les repeticions) que s’han situat en el pla mitjançant la tècnica t-SNE partint de la seva configuració en la darrera de les èpoques, la del Règim del 78. A continuació, els vectors que representen *article* en cada etapa s’han alineat amb el vector *article* en l’època del Règim del 78 fent ús de la llibreria `pwl` i s’han representat en el mateix pla. La figura 4 fa explícit el resultat aconseguit:

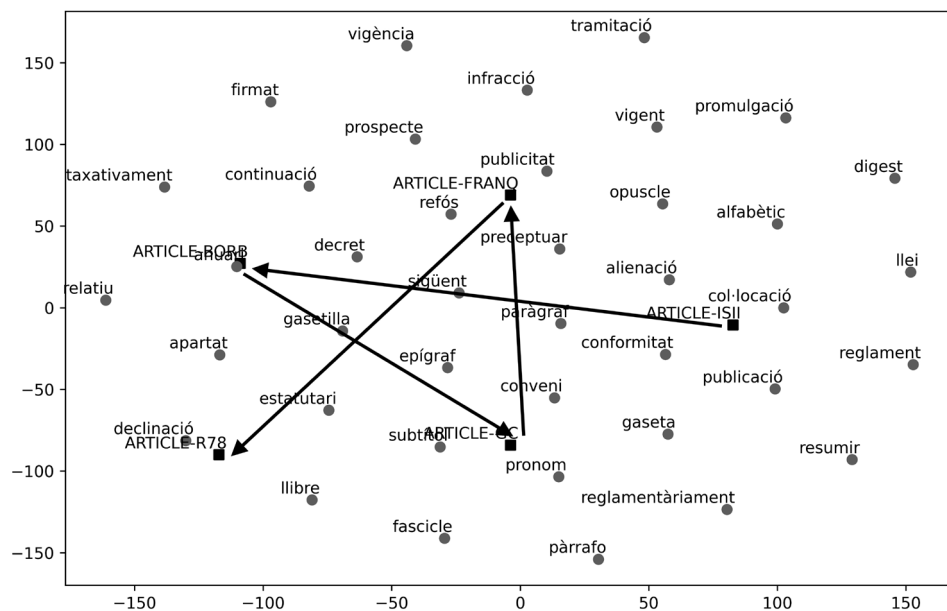


Figura 4. Evolució semàntica del mot *article*.

Tot i que la tècnica ha de ser encara refinada, sí que es pot apreciar una evolució del mot des del període isabelí fins al del Règim del 78. En general, el terme es desplaça des del context legal al literari, des de la referència a una part d'una llei a un entorn d'escriptura. Aquesta darrera inclusió és fins i tot més marcada durant la Guerra Civil que durant els anys que la segueixen, corresponents al Règim del 78.

5.2. Seguiment de cognats

Aquest mètode pot servir per efectuar un seguiment de cognats. El fet que el corpus encara estigui en construcció comporta que el nombre de paraules no sigui gaire elevat. Addicionalment, com que les dades no són en cap cas anteriors al segle XIX, si un dels elements del doblet ha entrat en desús, és possible que no se'n ressenyi cap entrada en èpoques tan tardanes. En el binomi *cultisme-patrimonial* és freqüent que aquest no tingui prou ocurrences per fer-ne un estudi aprofundit. Per aquesta raó, no és fàcil trobar doblets que es puguin tractar des de les dades disponibles en el CTILC. En la figura 5 es mostren els resultats del parell *càtedra-cadira*, dos mots que estan presents en totes les divisions temporals.

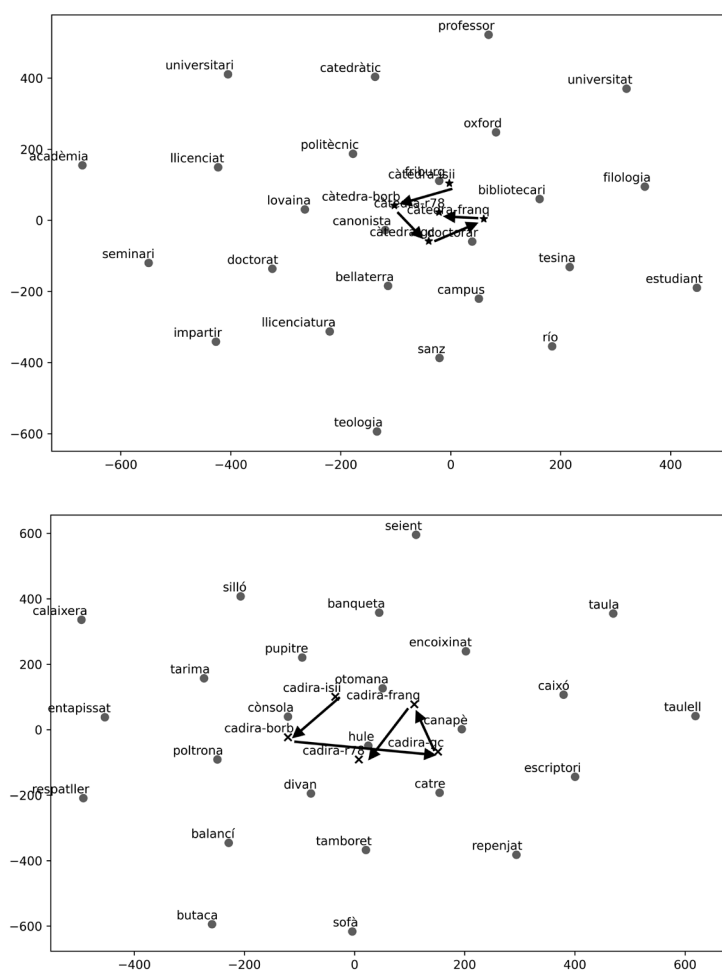


Figura 5. Evolució de *càtedra* (a dalt) i *cadira* (a baix).

Les imatges mostren que *càtedra* no s'ha mogut del seu espai semàntic durant els segles XIX i XX. En canvi, *cadira* ha experimentat un recorregut més ampli, tot i que dins del mateix camp semàntic. Aquestes dues evolucions es poden representar conjuntament en un mateix pla, com es pot veure en el gràfic de la figura 6.

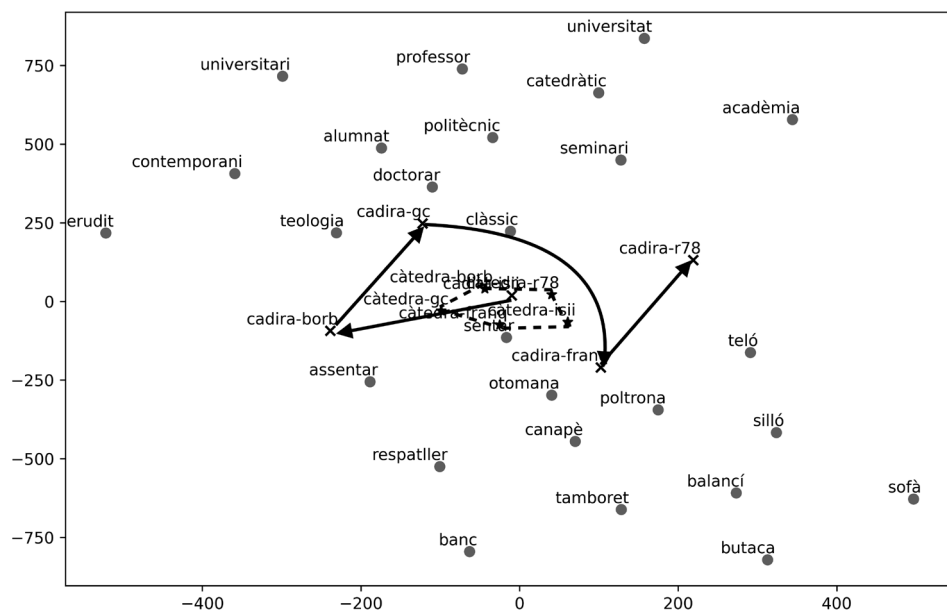


Figura 6. Evolució del parell *càtedra-cadira* dins el mateix pla.

6. CONCLUSIONS I TREBALL FUTUR

L'evolució semàntica amb mètodes computacionals és un àmbit d'estudi que tot just comença a ser abordat. Encara que en els darrers anys s'han fet algunes contribucions rellevants que obren el camí a les noves tècniques, cal encara aprofundir en els mètodes que poden resultar més adients per formular aquest tipus d'aproximació. A més, no hi ha un teixit teòric especialment dissenyat per a l'evolució semàntica que pugui donar un suport epistemològic als experiments que es duen a terme.

Un altre problema, més encara per a llengües amb relativament pocs recursos, com el català, és la manca de grans reculls de dades sobre els quals es puguin fonamentar els estudis de caire estadístic. En aquest sentit, iniciatives com la del CTILC poden ajudar a donar una gran empenta a la diacronia computacional.

Així, el futur d'aquest tipus d'estudis està encara per construir i depèn en bona mesura dels recursos existents. No obstant això, cal seguir buscant noves tècniques que puguin ajudar a entendre millor com es pot bastir una metodologia específica per a l'evolució semàntica basada en l'estudi quantitatiu de corpus.

REFERÈNCIES BIBLIOGRÀFIQUES

- BASTARDAS-BOADA, A. (2016). *Complexics as a Meta-Transdisciplinary Field*. www.pensamientocomplejo.org/blog/index.php/2017/02/lecturas-emergentes-del-congreso-mundial-de-pensamiento-complejo/.
- BOJANOWSKI, P.; GRAVE, E.; JOULIN, A.; MIKOLOV, T. (2016). *En-riching word vectors with subword information*. *Computing Research Repository*. arXiv:1607.04606 (2016). doi:10.1162/tacl_a_00051. <https://arxiv.org/abs/1607.04606>.
- CARRERAS, X.; CHAO, I.; PADRÓ, L.; PADRÓ, M. FreeLing (2004): *An Open-Source Suite of Language Analyzers Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC'04)*.
- CILLIERS *et al.* (2013). «Complexity, modeling, and natural resource management». *Ecology and Society*, vol. 18, núm. 3, pàg. 1.
- CLARK, S. (2015). «Vector space models of lexical meaning». A: LAPIN, S.; FOX, C. (eds.). *Handbook of contemporary semantic theory*. Oxford: Wiley-Blackwell, pàg. 493-522.
- DEVLIN, J.; CHANG, M. W.; LEE, K.; TOUTANOVA, K. (2018). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. arXiv:1810.04805v2.
- FIRTH, J. R. (1957). «A synopsis of linguistic theory 1930-1955». *Studies in Linguistic Analysis*, pàg. 1-32.
- GELL-MANN, M. (1995). «What is complexity?». *Complexity*, vol. 1, núm. 1, pàg. 16-19.
- GULORDAVA, K.; BARONI, M. (2011). «A distributional similarity approach to the detection of semantic change in the Google books ngram corpus». A: *Proceedings of the GEMS 2011 Workshop on GEometrical Models of Natural Language Semantics*. Edimburg: Association for Computational Linguistics, pàg. 67-71.
- HAMILTON, W. L.; LESKOVEC, J.; JURAFSKY, D. (2016). *Diachronic word embeddings reveal statistical laws of semantic change*. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Berlín, Germany: 1489-1501.
- HARRIS, Z. (1954). «Distributional structure». *Word*, vol. 10, núm. 23, pàg. 146-162.
- HOLLAND, J. H. (1992). «Complex adaptive systems». *Daedalus* [The MIT Press], vol. 121, núm. 1, pàg. 17-30.
- KULKARNI, V.; AL-RFOU, R.; PEROZZI, B.; SKIENA, S. (2015). «Statistically significant detection of linguistic change». A: *Proceedings of the 24th International Confe-*

- rence on World Wide Web, WWW'15, page 625–635, Republic and Canton of Geneva, CHE.
- MAATEN, L. J. P.; Van der HINTON, G. E. (2008). «Visualizing high-dimensional data using t-SNE». *Journal of Machine Learning Research*, núm. 9 (novembre), pàg. 2579-2605.
- MASSIP-BONET, À. (2018) (ed.). *Imbricacions entre variació històrica i variació geogràfica en la llengua catalana*. Barcelona: Edicions de la Universitat de Barcelona.
- MIKOLOV, Tomas (2013). «Distributed representations of words and phrases and their compositionality». *Advances in Neural Information Processing Systems*. <https://arxiv.org/abs/1310.4546>.
- PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V.; VANDERPLAS, J.; PASSOS, A.; COURNAPEAU, D.; BRUCHER, M.; PERROT, M.; DUCHESNAY, E. (2011). «Scikit-learn: machine learning in Python». *Journal of Machine Learning Research*, núm. 12, pàg. 2825-2830.
- PENNINGTON, J.; SOCHER, R.; MANNING, C. D. (2014). «GloVe: Global vectors for word representation». A: *Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, pàg. 1532-1543. doi:10.3115/v1/D14-1162.
- PETERS, M. E.; NEUMANN, M.; IYYER, M.; GARDNER, M.; CLARK, C.; LEE, K.; ZETTMAYER, L. (2018). *Deep contextualized word representations*. <https://arxiv.org/1802.05365>.
- VASWANI, A.; SHAZEER, N.; PARMAR, N.; USZKOREIT, J.; JONES, L.; GÓMEZ, A. N.; KAISER, L.; POLOSUKHIN, I. (2017). «Attention is all you need». 1st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA. <https://arxiv.org/abs/1706.03762>.

Modelitzacions matemàtiques, aprenentatge automàtic i ciència ciutadana per a l'estudi del canvi (socio)lingüístic

Ernest Querol Puig¹

Universitat Oberta de Catalunya

Emília López-Iñesta⁴

Universitat de València

1. INTRODUCCIÓ

Kabatek (2012: 32) fa un comentari molt encertat respecte a les introduccions dels articles que tracten el tema dels models matemàtics sobre la competició entre llengües:

Poderíase facer unha análise crítica do discurso dos artigos presentados aquí: todos comezan falando da extinción das linguas do mundo citando xeralmente as mesmas fontes, e todos espertan unha certa expectativa no lector cando falan da posibilidade de analizar obxectivamente as situacións de linguas en perigo ou das medidas terapéuticas para a salvación das mesmas.

Atès que hem observat el mateix que hi explica, no repetirem els tòpics que hi esmenta i farem una altra mena d'introducció. Serà molt menys general i se centrarà en el que pretenem en aquest article: des de principis del segle actual tant els físics com els matemàtics s'han llançat a proposar models aprofitant les eines de les seves disciplines en el camp de la sociologia del llenguatge.

En primer lloc explicarem breument què són els models i la modelització, en segon lloc presentarem de manera sintètica algunes modelitzacions dutes a terme en l'àmbit de la sociologia i en tercer lloc farem un repàs sumariíssim de les propostes fetes en la lingüística i en la sociolingüística variacionista, i esbossarem les nombroses modelitzacions dutes a terme en la sociologia del llenguatge, concretament en la competició entre llengües, àmbit en el qual comen-

¹ Aquest treball es va elaborar en el marc del projecte d'investigació de la Generalitat Valenciana GV/2021/110 i el projecte d'innovació de la Universitat de València UV-SFPIE PID20-1350001.

tarem tres investigacions del nostre Grup de Recerca Alcàntera: el Pont amb la Sociolingüística (GRA). Finalment, ressenyarem dues recerques més del GRA: una amb el recurs de l'aprenentatge automàtic (la classificació supervisada) i l'altra duta terme en l'àmbit de la ciència ciutadana.

2. ELS MODELS I LA MODELITZACIÓ

Des de l'epistemologia, Kremer-Marietti i Dhombres (2006: 11) assenyalen que els problemes que han de resoldre les ciències empíriques (les de la natura i les humanes) no són pas purament formals, sinó que avui cal resoldre el de l'experimentació, que es tendeix a descriure a partir de la noció de *model*, a la qual va associada la de *modelització*. La modelització té una doble cara: abstracta (l'àmbit teòric, del qual és una interpretació) i concreta (l'àmbit empíric, del qual és una formalització i una organització). La parella *model/modelització* té un valor representatiu del qual es pot explorar tant l'aspecte metodològic com l'epistemològic. Respecte a aquesta vessant epistemològica escriuen:

Modelitzar o fer una modelització és a partir d'un model trobar les expressions matemàtiques que representen esquemàticament i anàlogament el procés del fenomen. La fase de tria d'una modelització és pròpiament parlant una matematització. Però la modelització no és pas la simple traducció de les dades d'una disciplina a un altre llenguatge. Hi calen les matemàtiques per congelar el model en el sentit que no es pugui més subreptíciament cridar les propietats que seran afegides al model pres així com un punt de vista. Les matemàtiques són aleshores una eina epistemològica.

David W. Britt (1997: 4), en el seu llibre d'introducció als models, recull aportacions de diferents autors del darrer terç del segle xx. En primer lloc, diu que Richardson i Pugh (1981) donen una definició estàndard de *modelar* des de la perspectiva d'un sistema dinàmic: «El terme “model” significa una representació [...] d'alguna llesca de la realitat [...]. La finalitat és guanyar comprensió [...] exposant les assumpcions sobre un problema per a la crítica, l'experimentació i la reformulació».

En segon lloc, reporta la definició de model d'Uslaner (1976): «El model causal és una tècnica per seleccionar aquelles variables que són la [causa] potencial dels efectes —i [...] [aïllant] les contribucions separades als efectes fets per cada [causa]».

En tercer lloc, comenta una definició més general de Land (1969) que aporta les característiques típiques dels models: «[...] el terme *path model* es fa servir per a referir-se a un conjunt d'equacions estructurals [...] que representen el postulat causal i no causal [per exemple, merament associatiu] de les relacions entre les variables que es consideren».

El mateix Britt (1997: 14) afirma:

Els models ens forcen a ser més explícits i oberts respecte al que estem assumint, com emmarquem els problemes, tant els que estem estudiant com aquells que ignorem, i com pensem que unes coses estan relacionades amb unes altres. És evident que el lloc on estàs situat determina el que veus, però augmentant l'obertura i la capacitat de ser explícit del diàleg entre els elements essencials d'una situació, els models poden incrementar la riquesa i el rigor d'aquest diàleg.

I acaba l'epígraf dedicat a la fallibilitat del coneixement (Britt, 1997: 16) amb el paràgraf següent:

Les teories en les ciències socials poden ser en general vagues. Els models especifiquen temptativament quines variables creiem que són importants, el nostre pensament corrent sobre com són les seves naturaleses, i com pensem que estan relacionats amb altres variables del context. Els models són, per tant, més explícits que les teories. Faciliten la construcció del sentit fent explícites les implicacions alternatives de diferents teories, així incrementen la credibilitat de les implicacions restants.

Per cloure aquesta introducció, reportarem una representació esquemàtica de la seqüència de la modelització que fan Patrick Doreian i Norman P. Hummon (1976: 25).

La figura 1 mostra la seqüència de la modelització que s'inicia a partir de nocions vagues, considerades com una incursió crucial en la formulació del que s'ha d'estudiar. Una manera de convertir en coherents les nocions vagues és plantejar i respondre preguntes que construeixin teoria. Les qüestions metodològiques, teòriques i epistemològiques estan íntimament relacionades entre elles, fet que, en part, remet al procés. Així, el pròxim pas és que el model que s'ha de construir estigui lligat amb el procés i amb l'estructura. Una vegada articulat un model, s'ha d'expressar de manera matemàtica, és a dir, en un sistema d'equacions. Els paràmetres d'aquest model són els paràmetres d'interès i molts d'ells no es poden calcular directament. S'ha de construir un control estructural del model en la forma d'un sistema simultani amb equacions diferencials lineals de primer ordre. Els paràmetres d'aquest sistema d'equacions

no són calculables directament. Cal una teoria de l'estimació i, una vegada obtinguda, podrem establir les dades requerides per al procediment d'estimació per a ser realitzat, és a dir, per a ser descrit empíricament.

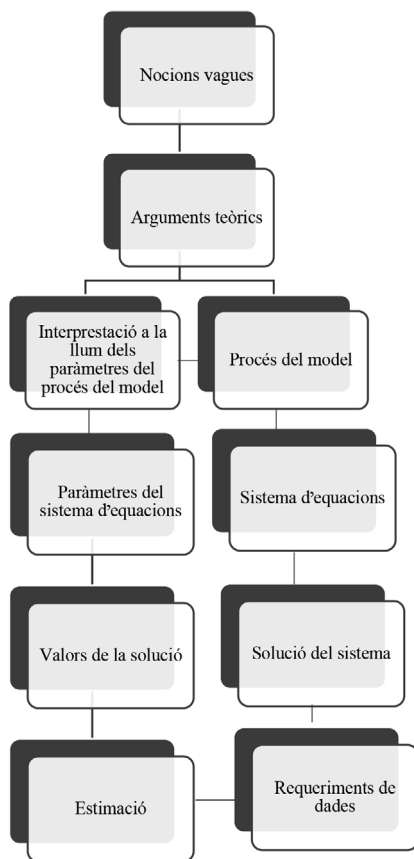


Figura 1. Esquema de la seqüència de modelització

Font: Patrick Doreian i Normal P. Hummon (1976: 25).

I, justament respecte a les dades, Scott de Marchi (2005: 171) escriu: «Especificant les explicacions en competició i testant-les amb les dades és la millor manera de construir la teoria. I quan les dades són menys que perfectes, les implicacions lògiques poden servir com un substitut excel·lent».

Tal com hem anunciat, abans d'analitzar les modelitzacions en els àmbits de la lingüística, de la sociolingüística variacionista i de la sociologia del llenguatge, comentarem breument les proposades en sociologia.

2. LES MODELITZACIONS EN SOCIOLOGIA

Esbossarem una síntesi de les modelitzacions en sociologia feta per Thomas F. Fararo (2011: 99-120). Hi presenta el procés generatiu de la construcció de models i comença pel que anomena «teorització analítica en la perspectiva postclàssica». Afirmar que el període que va des de la dècada de 1930 fins a la de 1970 va ser l'època en què la cerca d'un marc teòric general per a l'anàlisi dels fenòmens socials va arribar al seu punt àlgid. Va ser justament al Harvard de mitjan anys trenta quan algunes influències intel·lectuals van ajudar a fomentar aquesta aspiració ambiciosa de Talcott Parsons i George Homans. El llibre d'Alfred North Whitehead *Science and the modern world*, publicat el 1925, subratllava la importància d'un esquema conceptual per a la ciència. El 1950 Homans va començar a explicitar els termes per a un esquema conceptual de la teoria dels grups humans com un sistema social. Un any després, Parsons feia aportacions en aquesta mateixa línia i se centrava en el problema que Whitehead va anomenar «la fallàcia de la concreció fora de lloc».

L'esquema de Homans incloïa tres categories: la freqüència de la interacció, la intensitat del sentiment positiu i la semblança de l'activitat. Deixava clar que la forma ideal d'una teoria analítica d'un sistema social dinàmic era un sistema d'equacions diferencials. Aquesta idea té la màxima credibilitat en física: Margenau (1950) arriba a una conclusió semblant després de fer una anàlisi basada en les impressions de les representacions formals emprades en camps molt diversos de la física. Tanmateix, aquestes aportacions de Homans no van ser seguides per altres investigadors.

Parsons va fer un bon nombre d'aportacions, però no pretenia la transició a una teoria analítica com un sistema de lleis analítiques com havia suggerit en la seva agenda en *The structure of social action* (1937). El problema era que les seves explicacions teòriques no tenien generativitat. Fararo (2011: 118) escriu que aquest concepte es refereix al comportament social dels actors i, en particular, a com la concatenació (interacció o interdependència) de les seves accions genera resultats col·lectius. Relaciona la generativitat amb els models matemàtics i fa un repàs de la seva aplicació en sociologia segons la forma que adopten: estocàstica (probabilística) o determinista. Finalment, se centra en la generativitat dels models cibernètics i de la sociologia computacional.

3. LES MODELITZACIONS EN LINGÜÍSTICA I EN SOCIOLINGÜÍSTICA VARIACIONISTA

Una de les primeres modelitzacions que coneixem en el camp de la lingüística és la d'Herdina i Jessner (2002). Proposen un model dinàmic del multilingüisme des de la psicolingüística. Relacionen la teoria de l'adquisició d'una segona llengua amb les recerques sobre el bilingüisme i sobre la teoria dels sistemes dinàmics. Analitzen els canvis de la competència lingüística multilingüe en un nivell individual per proporcionar una descripció holística del multilingüisme.

Un intent semblant, però molt més teòric, és el que presenten Verspoor, De Bot i Lowie (2011). Pretenen proposar un model dinàmic de creixement lingüístic referit a la segona llengua. Conclouen que les dades no s'ajusten prou al model i que caldrà introduir-hi un retard en la retroacció. Hi assenyalen les possibilitats que la modelització dinàmica els ofereix per testar hipòtesis.

William A. Kretzschmar (2015: 56-80) analitza el llenguatge i els sistemes complexos i fa un repàs de les aportacions de diferents autors des dels anys vuitanta. Comenta les propostes de Paul Hopper (1987) sobre la gramàtica emergent. Aquest darrer lingüista, acompanyat d'Elizabeth Traugott (1993), també en fa sobre la gramaticalització. Joan Bybee estudia la fonologia i els usos lingüístics. Janet Pierrehumbert (2001) tracta el problema de la gestió de la freqüència d'informació i distingeix clarament la teoria exemplar dels models «estàndard» de fonètica i fonologia. Adèle Goldberg (2006) defensa que tots els nivells d'anàlisi gramatical impliquen construccions: emparellaments apresos de la forma amb la semàntica o amb la funció del discurs, incloent-hi morfemes o paraules, modismes parcialment lexicalitzats, patrons generals de frase plens i complets. Un altre punt essencial de la seva aportació és que les construccions es combinen lliurement amb les formes actuals d'expressions sempre que no estiguin en conflicte. Michel Tomassello (2003) concep la gramàtica com un inventari de construccions. Nick Ellis i Diane Larsen-Freeman (2009) estudien el llenguatge com un sistema complex adaptatiu. I, finalment, Joan Bybee (2010) en el seu llibre *Language, usage and cognition* té la concepció més aprofundida del llenguatge com un sistema complex adaptatiu.

La concepció de la gramàtica com un sistema complex ofereix un mitjà per distingir la gramàtica tradicional prescriptiva de les gramàtiques científiques contemporànies de maneres diferents. Fins i tot podria relaxar la competició, de vegades estrident, entre els promotors dels diferents tipus de gramàtiques. Kretzschmar (2015: 104) conclou el capítol que dedica a l'enfocament sistèmic amb el paràgraf següent:

Una vegada compreguem la relació de la gramàtica amb els sistemes complexos de la parla, serem lliures de realitzar una feina millor fent el tipus de generalitzacions gramaticals que sempre hem volgut fer sobre el llenguatge, i ens cal fer-les per tal d'explotar el que Saussure reconeixia com un valor lingüístic. La gramàtica *per se* no és el resultat del sistema complex de la parla, però no podem fer una gramàtica satisfactòria sense comprendre el sistema complex de la parla.

Els sistemes complexos també poden ser molt útils per a l'estudi de la història de les llengües. Kretschmar (2015: 130) afirma que els lingüistes actuals han d'adoptar la millor ciència disponible per guiar el seu treball, que, a hores d'ara, és la dels sistemes complexos.

Un dels grans avantatges de fer servir la ciència de la complexitat és que podem connectar el nostre camp de recerca (la lingüística, la sociolingüística variacionista o la sociologia del llenguatge) amb molts altres camps que estudien els sistemes complexos, i adoptar-hi els principis i els mètodes que han estat satisfactoris en altres disciplines. Passem ara a veure els avantatges que pot proporcionar a la sociolingüística variacionista concebre la investigació en el marc dels sistemes complexos. Tinguem presents els cinc principis bàsics que els caracteritzen, segons Kretschmar (2015: 193):

- (1) Activitat dinàmica contínua en el sistema = converses i escriptura
- (2) Interacció aleatòria d'un gran nombre de components = realitzacions aleatòries de les característiques de les variants
- (3) Intercanvi d'informació amb *feedback* = comparació dels components usats per diferents parlants en diferents contextos
- (4) Reforç dels comportaments = probabilitat de canvi dels components en usos futurs
- (5) Emergència de patrons estables sense un control central = varietats regionals/socials/textuals

Kretschmar (2015: 200) clou aquest capítol dient que la sociolingüística variacionista també encaixa molt bé en la nova ciència dels sistemes complexos. Coincideix amb els plantejaments fundacionals de la disciplina perquè la ciència de la complexitat prediu els tipus de regles que es donen en la parla i també prediu la immensa varietat de possibles comunitats de parla que emergeixen i coexisteixen en diferents nivells i que continuen canviant al llarg del

temps, que justament són l'àmplia gamma de varietats que va atreure els variacionistes de bon començament. En aquest sentit, és possible reanalitzar les dades obtingudes fins ara, per a la qual cosa seria desitjable fer públiques les bases de dades que té cada investigador. De tota manera, sense dades també caldria visitar els llibres i els articles per trobar les maneres de resoldre els problemes del maneig de les dades i les anàlisis originals dels investigadors. S'haurien de rebutjar les dades que no complissin la distribució i les propietats d'escala dels sistemes complexos. I acaba el seu llibre amb les paraules següents (Kretzschmar, 2015: 212):

La complexitat és l'estil de l'explicació que millor encaixa amb una aproximació científica a la nostra matèria i als problemes essencials. Si els nous humanistes poden localitzar i descriure els nostres propis fets dins de la «forma científica» dels sistemes complexos, tenim una oportunitat millor que les hem tingut fins ara de descriure les interaccions humanes que han estat sempre del nostre interès.

4. LES MODELITZACIONS EN LA SOCIOLOGIA DEL LLenguatge

The splendid simplifications which made life easier for computational linguists in the 70s and 80s are not helpful in the realm of sociolinguistics and language contact studies.

WOLFGANG WILDGEN

Presentarem un breu estat de la qüestió de les modelitzacions proposades en la sociologia del llenguatge, concretament les referides als processos de substitució, manteniment i capgirament lingüístic, que en la bibliografia escrita en anglès se sol anomenar «competició entre llengües».²

Les analogies en què s'han basat les modelitzacions dels processos lingüístics han provingut bàsicament de dos paradigmes: del de la difusió social i del de depredador-presa. Altres enfocaments com la formació de l'opinió, les dinàmiques de la població, la lluita entre ideologies i l'expansió de les religions i de les esglésies també han atret l'interès dels físics per aplicar-hi la mecànica estadística i la teoria dels sistemes complexos.

² Michel Boissonneault i Paul Vogt (2021) fan un estat de la qüestió molt complet dels articles publicats en anglès sobre els models matemàtics aplicats a la competició de llengües; per raons d'espai no podem comentar les seves aportacions.

Per començar, hem d'esmentar els treballs de Freedman (1980) sobre models matemàtics deterministes en la població ecològica, en què explica molt bé la versatilitat d'aquest paradigma. Deu anys més tard, juntament amb Baggs (Baggs i Freedman, 1990) trien la metodologia del depredador i de la presa per estudiar els processos de substitució i manteniment de les llengües, que ells anomenen «persistence» i «extinction». Les matemàtiques que fonamenten aquesta modelització són les que el 1925 Alfred Lotka i un any més tard Vito Volterra van proposar. El creixement de la població és malthusià, és a dir, la població creix exponencialment, mentre que l'aportació del menjar s'espera que creixi de manera aritmètica. Baggs i Freedman continuen publicant en aquesta mateixa línia dos articles més el 1991 i el 1993 sobre l'equilibri i la supervivència de les llengües.

Com hem avançat, els paradigmes de la difusió i l'epidemiològic també pretenen modelitzar la difusió social. Aquest darrer paradigma fa servir equacions diferencials ordinàries per modelitzar l'expansió d'una malaltia en la població. El paradigma de la difusió és una extensió de l'epidemiològic i s'empra per analitzar l'augment de popularitat d'un fenomen i, especialment, les vendes d'un producte comercial. Una aplicació pionera de la distribució de la innovació tecnològica va ser la de Coleman (1964) respecte a l'expansió de l'ús de les tècniques mèdiques.

Aquest model de difusió epidemiològica s'ha fet servir també per a fenòmens més abstractes, com ara l'expansió de les esglésies i també per entendre l'expansió dels disturbis.

Un altre intent primerenc d'estudiar la substitució lingüística des de l'àmbit de l'ecologia va ser el de T. J. Case el 1999. Hi considera l'escenari més simple de competició entre dues espècies, mitjançant les clàssiques equacions de Lotka-Volterra. Altres recercadors que fan servir aquestes equacions són Mira i Paredes (2005), Pinasco i Romanelli (2006), Wyburn i Hayward (2008), Mira, Seoane i Nieto (2011) i Zhang i Gong (2013).

Un altre tipus d'enfocament és el que empra les equacions diferencials. Només amb una equació diferencial simple, Abrams i Strogatz (2003), i amb només una pàgina publicada a *Nature*, van desencadenar una cascada de rèpliques i d'intents de millora considerable. Aquests matemàtics van estudiar el declivi de quaranta-dues llengües minoritàries mitjançant una simple equació diferencial que recull la probabilitat que un parlant vulgui canviar la seva llengua en funció, en primer lloc, de la quantitat de parlants de l'altra llengua i, en segon lloc, del prestigi percebut.³ Aquesta simplicitat comporta moltes restriccions al

3 Per a un bon resum de la proposta d'Abrams i Strogatz, consulteu Richard A. Blythe (2015: 5).

model perquè no permet la solució d'una població bilingüe estable, ja que només considera el monolingüisme dels dos grups lingüístics en lluita. Malgrat tot, com hem avançat, aquest article va ser l'origen d'un llarg seguit de propostes per millorar-lo.

De seguida, Patriarca i Leppanen (2005) van afegir al model d'Abrams i Strogatz el paràmetre de la difusivitat, que optava perquè el prestigi de cada llengua fos diferent en les distintes zones, fet que permetia que en les simulacions hi hagués períodes en què les dues llengües sobrevisquessin malgrat que tinguessin valors de prestigi diferents.

Per superar les limitacions del model d'Abrams i Strogatz, Mira i Paredes (2005) van introduir els bilingües en la seva proposta i, també, la semblança de les llengües en contacte; aquest segon paràmetre, però, ha estat molt criticat per la seva manca de fonamentació.

Aquestes modelitzacions tenen un enfocament macroscòpic, a diferència de les que recorren als models basats en agents, que el tenen microscòpic. Una simulació d'aquest darrer tipus és un model en el qual elements discrets (anomenats «agents») representen les entitats o grups d'entitats seleccionats del sistema que s'està modelant. Les dinàmiques d'un model basat en agents estan descrites localment en termes de com els agents individuals interactuen entre ells. Wang i Minett (2005) fan servir els models basats en agents les simulacions dels quals mostren fenòmens interessants com ara l'autoorganització de les poblacions en l'espai que depèn de l'estructura de les xarxes socials. En la seva proposta també consideren els bilingües i la taxa de mortalitat. Altres autors que fan servir els models basats en agents⁴ són Kosmidis *et al.* (2005), que empen les simulacions Monte Carlo, basades en la teoria de jocs, com també ho fan Castelló i els seus col·legues de la Universitat de les Illes Balears: Castelló *et al.* (2007), Toivonen *et al.* (2009) i Vázquez *et al.* (2010). I des de la Universitat de Barcelona també Beltran *et al.* (2009) mostren que l'aplicació dels autòmats cel·lulars a l'estudi dels processos de substitució lingüística és un mètode amb el qual es pot investigar l'estructura interna del grup bilingüe.⁵

Castelló, Eguíluz i San Miguel (2006) conclouen que la consideració dels bilingües pot tenir com a resultat la coexistència de les llengües.

4 Macy i Willer (2002 : 162) assenyalen que els models basats en agents fan servir la simulació per trobar mecanismes causals que poden subratllar les associacions estadístiques.

5 Tal com assenyalen Castellano *et al.* (2008), hi ha dos nivells de dificultat en la investigació dels models de la dinàmica social. La primera dificultat rau en la definició raonable i real dels models microscòpics, i la segona, en el problema habitual d'inferir la fenomenologia macroscòpica fora de la dinàmica microscòpica d'aquests models. Obtenir resultats útils d'aquests pot semblar una tasca desesperant.

També tenen una mirada microscòpica les diferents aportacions fetes des del camp de la teoria econòmica. Thorstein Templin (2019: 41) assenyala que els parlants pretenen maximitzar la seva utilitat individual quan prenen decisions relacionades amb la llengua. François Grin (1992) va fer una de les primeres propostes aplicades a un model lingüístic dinàmic. Posteriorment, John i Yi (2001) presenten una versió d'un model amb dos períodes amb generacions successives. El mateix Templin (2019: 43-44) no segueix les modelitzacions fetes des de la física i les matemàtiques, sinó que també se centra en la perspectiva econòmica. Conceptualitza les famílies i els individus adults com a agents que maximitzen la utilitat; aquesta aproximació l'ajuda a extreure variables generals en els processos lingüístics de transmissió i d'adquisició.

Un altre enfocament de l'evolució dels models de competició entre llengües és el que es fa des de la perspectiva del problema del consens. Ryo Fujie, Kazuyuki Aihara i Naoki Masuda (2013) el plantegen respecte a les opinions. Es demanen si entre diferents estats de competició s'aconsegueix la unanimitat d'un d'ells. Diferents disciplines com la ciència política, la sociologia i la física estadística estudien aquests mecanismes. En els models de formació del consens s'assumeix que cada individu posseeix un dels diferents estats que pot variar amb el temps. L'índex de variació depèn del context, com, per exemple, el nombre d'individus que adopta un estat diferent. La física estadística ha abordat aquest problema analitzant una varietat de models, com ara el del vot, el de la norma de la majoria, el de la confiança acotada, el d'Axelrod i el *naming game*.

El mecanisme que podria portar al consens en una població és la preferència de la majoria. La formació de l'opinió col·lectiva sota diferents majories de patrons de vot ha estat examinada per a poblacions mitjanes i en diferents tipus de xarxes, com les gelosies regulars i amb xarxes de petit abast, xarxes heterogènies i xarxes amb estructura comunitària.

La preferència de la majoria pot ser identificada amb l'aversion a la minoria. Quan només hi ha dos estats, aquests són equivalents perquè un estat és la majoria si i només si l'altre estat és la minoria, que és el que es donava en el model d'Abrams i Strogatz. Tanmateix, els dos principis poden ser distints si hi pot haver més de dos estats; així, dues llengües poden coexistir segregades espacialment en un model en el qual la dinàmica de competició i la difusió espacial estan combinades, tal com proposen Patriarca i Leppänen (2004) i Patriarca i Heinsalu (2009).

Chapel *et al.* (2010) afirmen que la coexistència també es dona quan l'atracció de les llengües es manipula dinàmicament o quan el bilingüisme, horitzontal (d'adults a adults) i vertical (d'adults a nens), la transmissió de les llengües

i el dinamisme de l'atracció de les llengües es combinen, com estudien Minett i Wang (2008) i Wang i Minett (2005).

Sutantawibul *et al.* (2018) interpreten els censos lingüístics i prediuen la substitució lingüística del català a Catalunya, del castellà a Huston (Texas), del neerlandès a Brusselles, de l'euskera al País Basc i del francès al Canadà. Apliquen tres models matemàtics: el d'Abrams-Strogatz, el de Castelló i el de Mira. Conclouen que el model d'Abrams-Strogatz no pot donar compte de totes les evolucions i que en el de Mira cal millorar com tracta les diferències entre llengües.

5. LES APORTACIONS DEL GRUP DE RECERCA ALCÀNTERA: EL PONT AMB LA SOCIOLINGÜÍSTICA (GRA)

5.1. Les aportacions respecte a la competició entre llengües

Com hem presentat en el punt 2, a mitjan segle xx tant des de la sociologia com des de la física s'arribava a la conclusió que la forma ideal d'una teoria analítica d'un sistema social dinàmic era un sistema d'equacions diferencials.

Les tres recerques que ressenyarem —elaborades pel nostre grup de recerca Alcàntera: el Pont amb la Sociolingüística (GRA)— s'inscriuen en l'anàlisi dels processos de substitució i de reversió lingüística. La modelització del sistema mitjançant equacions diferencials ens permet estudiar si la dinàmica del sistema és de desaparició o no. En el cas que no fos una dinàmica de desaparició, serien necessaris nous models que permetessin canvis de tendència. Aquesta modelització és una eina fonamental per a l'estudi de sistemes socials.

La primera aportació que hem fet ha estat Miralles, Perucho i Querol (2015), que és una aplicació del model dinàmic de competició entre llengües d'Abrams i Strogatz al cas català-castellà al País Valencià. Hi presentem una recerca inicial de la dinàmica de l'evolució del procés de substitució lingüística al País Valencià a partir del seu model matemàtic d'equacions diferencials. En particular, discutim l'aplicació del model esmentat a dades empíriques obtingudes d'enquestes realitzades (els anys 1989, 1992, 1995, 2005 i 2010) pel Servei d'Investigació i Estudis Sociolingüístics (SIES) de la Generalitat Valenciana i per l'Acadèmia Valenciana de la Llengua (AVL) l'any 2004. En el cas de les enquestes de l'AVL, hem dut a terme una transformació de les dades obtingudes per franges d'edat a una distribució en el temps.

Partint dels ajusts dels paràmetres que controlen els models, es presenta un estudi de la dependència de l'evolució de l'ús del valencià amb aquests parà-

metres. S'han pogut establir uns certs rangs de valors possibles dels paràmetres que descriuen la interacció entre les dues llengües. Així, el model d'Abrams i Strogatz, dins de la seva limitació per descriure un sistema com el tractat en aquesta investigació, ens ha servit per fer-nos una idea de l'interval de variació dels paràmetres, que hem emprat posteriorment en el model de Mira-Paredes. Com hem vist en l'apartat 4, aquests dos físics plantegen un model més general per tal de refinar el d'Abrams-Strogatz. Tal com hem avançat, incorporen en el model un grup de parlants bilingües i hi introdueixen un nou paràmetre que representa la probabilitat que un parlant monolingüe no passi a l'altra llengua, sinó que adopti una pràctica bilingüe. Els autors identifiquen aquest paràmetre amb la semblança entre les dues llengües. El problema fonamental a què ens vam enfrontar durant l'estudi va ser la manca de dades, que va fer que no es poguessin trobar valors numèrics concrets dels paràmetres, sinó més aviat intervals dins dels quals aquests poden variar. Així, l'aplicació dels models ens va permetre quantificar, tot i que amb incerteses considerables, les variables que modelen la interrelació entre el català i el castellà i trobar conjunts que descriuen apropiadament les dades de les enquestes.

La segona aportació ha estat el treball de final de grau, encara inèdit, de Maitte Gandia (2019) *Sistemes dinàmics tipus Lotka-Volterra: una aplicació social*, també dirigit (com l'anterior) per Manel Perucho. Els objectius de la recerca eren:

- (1) Estudiar un sistema dinàmic tipus Lotka-Volterra amb les modificacions introduïdes per Abrams i Strogatz.
- (2) Modificar les equacions per estudiar com canvia l'evolució del sistema en funció d'un dels paràmetres: l'estatus.
- (3) Generalitzar el model de mort lingüística d'Abrams i Strogatz a un model en què coexisteixin dues llengües.

S'hi van dur a bon port els dos primers punts; el tercer, però, només va mostrar períodes efímers de coexistència.

La tercera aportació és la de Miralles *et al.* (2017) *La modelització matemàtica en la sociologia del llenguatge: estudi comparatiu de l'evolució dels usos lingüístics a Catalunya i al País Valencià*.

La recerca volia respondre la pregunta següent: és aplicable el model dinàmic de la desaparició de les llengües proposat el 2003 per Abrams i Strogatz a les dades demolingüístiques del País Valencià i de Catalunya de què disposem?

En l'article concloem que el model d'Abrams i Strogatz és aplicable a la tendència decreixent del nombre de parlants de català que es dona al País Va-

lencià, però, per contra, no és aplicable a l'evolució del sistema a Catalunya. Aquest resultat va en la línia de l'aconseguit per Gandia (2019).

5.2. Les aportacions respecte als prejudicis lingüístics mitjançant la ciència ciutadana

La ciència ciutadana és un bon mecanisme per fer recerca en l'àmbit de la sociologia del llenguatge, ja que proposa la participació i l'arreglada de dades del gran públic, alhora que genera una cultura científica en la societat. El GRA ha col·laborat amb el projecte Milmots, un projecte de ciència ciutadana, iniciativa de la Universitat de València, que té l'objectiu de difondre, conèixer i compartir el nostre lèxic, les nostres paraules associades als nostres pobles i a les nostres comarques.⁶ Però què és la ciència ciutadana?

El terme *ciència ciutadana* va ser encunyat el 1995 per Alan Irwin per descriure la col·laboració entre la ciutadania i les persones dedicades professionalment a la ciència (Curtis, 2018). De fet, el 2002 Irwin assenyalava: «La ciència ciutadana hauria de respondre a les inquietuds i necessitats dels ciutadans, reconeixent i validant el coneixement científic que la ciutadania produeix».

Tal com indica l'Oficina de Ciència Ciutadana de l'Ajuntament de Barcelona, s'entén per *ciència ciutadana* la recerca científica que té la implicació activa del públic no especialitzat.⁷ En aquest sentit, el *Llibre blanc de la ciència ciutadana a Europa* (Sanz *et al.*, 2021) associa aquest terme a la participació del públic en general en activitats d'investigació científica, que contribueix activament a la ciència amb el seu esforç intel·lectual, el seu coneixement circumdant o les seues eines i recursos.

Els beneficis de la ciència ciutadana també es poden traslladar a l'àmbit de l'educació i permeten incorporar als currículums educatius diferents pràctiques que es poden implementar tant a l'aula com a espais públics com ara les biblioteques (López-Iñesta *et al.*, 2021). El projecte Ciència Ciutadana a les Biblioteques, de la Diputació de Barcelona, ofereix una visió sobre com les biblioteques públiques poden establir una poderosa sinergia amb la ciència ciutadana garantint l'accés universal al coneixement.⁸ Actualment, a més, tenen un gran potencial, per exemple, de convertir-se en espais de generació de coneixement vàlid a escala local, però que pot ser degudament articulats a escala global.

6 <https://milmots.eu>.

7 www.barcelona.cat/barcelonaciencia/ca/ciencia-ciutadana.

8 <https://cutt.ly/jWj9uAz>.

Amb la idea de veure les biblioteques públiques com a espais on generar nou coneixement amb i per a la ciutadania, la Gerència de Serveis de Biblioteques Municipals de la Diputació de Barcelona, mitjançant el programa BiblioLab, va impulsar el Laboratori de Ciència Ciutadana, un curs de formació impartit per OpenSystems, de la Universitat de Barcelona, i adreçat als i les professionals d'aquests equipaments.⁹

En el Laboratori, les biblioteques van seleccionar deu projectes de ciència ciutadana que estan actualment en funcionament i van imaginar com seria la implantació del projecte al seu municipi, tot adaptant-lo a les particularitats culturals i socioambientals del seu espai, barri i població.

La iniciativa Milmots va ser triada com un dels recursos seleccionats pel projecte Ciència Ciutadana a les Biblioteques per la seua relació amb la sociolingüística.

Milmots va néixer amb la finalitat de conèixer i difondre el nostre lèxic a través de la recollida de les paraules i les locucions aportades voluntàriament pels internautes que es localitzaven en un mapa, com es pot veure en la figura 2, on s'hi podien consultar i ordenar els resultats.

Les aportacions dels participants anònims de Milmots pel que fa a les seues tries lèxiques proporcionen la matèria primera per a la recerca científica, la qual és compatible i complementària amb el vessant lúdic del web. Sobre aquesta base, els participants d'aquest web interactiu han aportat prop de quinze mil mots i locucions dels diferents territoris on es parlen varietats dialectals de la nostra llengua, tant del País Valencià, Catalunya, les Illes Balears, la Franja de Ponent, la Catalunya del Nord o l'Alguer, com d'altres territoris amb llengües diferents.

L'article de Grimaldo *et al.* (2017) posa de manifest que l'elevat volum de mots i locucions recollit ha proporcionat un material idoni per fer un estudi sociolingüístic que, mitjançant l'aplicació de procediments estadístics que en garanteixen la fiabilitat, posa en relació les variables socials dels participants (edat, àrea dialectal i nombre d'habitants de la població) amb variables lingüístiques derivades principalment de la normativitat dels mots i de les locucions aportats.

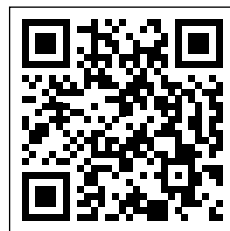


Figura 2. Imatge del web de Milmots.

Font: elaboració pròpia a partir de la pàgina web del projecte Milmots (<https://milmots.eu/mapa.php>).

⁹ <https://ub.edu/opensystems/ca>.

En el context de la ciència ciutadana, els resultats obtinguts se separen dels prejudicis que atribueixen un coneixement superior de la normativa lingüística a la població més jove, als residents de l'àrea dialectal central i als habitants de localitats més poblades. Alhora, els resultats s'alineen amb el marc teòric exposat en el treball amb relació a l'edat, a l'àrea dialectal i al nombre d'habitants de les poblacions d'on provenen els participants.

6. CONCLUSIÓ

En les dues darreres dècades l'estudi del canvi (socio)lingüístic s'ha dut a terme des de diferents perspectives. Aquest article ha esbossat breument les modelitzacions fetes en l'àmbit de la sociologia, de la lingüística i de la sociolingüística variacionista, i s'ha centrat en les proposades en la sociologia del llenguatge, concretament en l'estudi dels processos de substitució i reversió lingüística. S'hi han presentat els resultats de diferents investigacions del GRA que tracten tant la modelització del procés de substitució lingüística com la predicció del canvi sociolingüístic (fent ús de tècniques d'intel·ligència artificial) i també l'estudi dels prejudicis lingüístics mitjançant la ciència ciutadana.

En primer lloc, s'han ressenyat tres recerques del GRA sobre els processos de substitució i de reversió lingüística dutes a terme mitjançant dos tipus diferents de modelització matemàtica. En la primera s'ha aplicat l'equació diferencial d'Abrams i Strogatz a dades sociolingüístiques del País Valencià. En la segona s'han fet servir les equacions del tipus Lotka-Volterra aplicades també a dades de parlants d'aquest mateix territori. I en la tercera s'ha comprovat si també es podia validar el model d'Abrams i Strogatz per a dades de parlants de Catalunya, cosa que no ha estat possible perquè l'evolució dels usos lingüístics d'aquest territori no va cap a l'extinció.

En segon lloc, s'ha mostrat una aplicació de tècniques d'aprenentatge automàtic per fer una classificació supervisada mitjançant la implementació de mètodes computacionals que permeten induir models de coneixement a partir d'informació que prové de dades d'exemple disponibles. Els mètodes de classificació supervisada permeten entrenar models de predicció del grau d'ús del català amb un percentatge d'encert que supera els aconseguits en les investigacions precedents.

Finalment, s'ha presentat el projecte Milmots, una eina interactiva que permet arreplegar paraules i locucions, visualitzar-les en un mapa i plantejar jocs per conèixer i difondre el lèxic mitjançant un projecte de ciència ciutadana

amb una vessant lúdica i amb components típics de la gamificació. El disseny d'eines com Milmots proporciona un material idoni per fer un estudi sociolingüístic que permet posar en relació diferents variables socials dels participants, com ara l'edat, l'àrea dialectal i el nombre d'habitants de la població on viuen, amb variables lingüístiques derivades principalment de la normativitat dels mots i de les locucions aportats.

En definitiva, amb les investigacions del GRA que hem presentat hem pretès mostrar que hi ha altres maneres d'abordar els estudis sociolingüístics: els models matemàtics, les tècniques d'aprenentatge automàtic i la ciència ciutadana.

REFERENCIAS BIBLIOGRÁFICAS

- ABRAMS, D. M.; STROGATZ, S. H. (2003). «Modelling the dynamics of language death». *Nature*, núm. 424, pàg. 900.
- BAGGS, I.; FREEDMAN, H. I. (1990). «A mathematical model for the dynamics of interactions between a unilingual and a bilingual population: Persistence versus extinction». *Journal of Mathematical Sociology*, vol. 16, núm. 1, pàg. 51-75.
- BAGGS, I.; FREEDMAN, H. I. (1991). «Convergence to equilibria in general models of unilingual-bilingual interactions». A: BUSENBERG, S. et al. (eds.). *Differential equations models in biology, epidemiology and ecology*. Berlín, Heidelberg: Springer.
- (1993). «Can the speakers of a dominated language survive as unilinguals?: a mathematical model of bilingualism». *Mathematical Computational Modelling*, vol. 18, núm. 6, pàg. 9-18.
- BELTRAN, F. S.; HERRANDO, S.; FERRERES, D.; ESTREDER, V.; ADELL, M. A.; RUIZ-SOLER, M. (2009). «Forecasting language shift based on cellular automata». *Journal of Artificial Societies and Social Simulation*, vol. 12, núm. 3, pàg. 5. www.jasss.org/12/3/5.html.
- BELTRAN, F. S.; HERRANDO, S.; FERRERES, D.; ESTREDER, V.; ADELL, M. A.; RUIZ-SOLER, M. (2011). «A language shift simulation based on cellular automata». A: BLANCHARD, E. G.; ALLARD, D. *Handbook of research on culturally-aware information technology: Perspectives and models*. Nova York: Hershey.
- BLYTHE, R. A. (2015). «Colloquium: Hierarchy of Scales in Language Dynamics». *The European Physical Journal B*, núm. 88, pàg. 295.
- BOISSONNEAULT, M.; VOGT, P. (2021). «A systematic and interdisciplinary review of mathematical models of language competition». *Humanities and Social Science Communications*, núm. 8, pàg. 21.
- BONNEY, R.; PHILLIPS, T. B.; BALLARD, H. L.; ENCK, J. W. (2016). «Can citizen science enhance public understanding of science?». *Public Understanding of Science*, vol. 25, 1, pàg. 2-16.

- BRITT, D. W. (1997). *A conceptual introduction to modelling: Qualitative and quantitative perspectives*. Mahwah: Lawrence Erlbaum Associates.
- BUSENBERG, S.; MARTELLI, M. (eds.) (2013). *Lecture notes in biomathematics*. Vol. 92, *Differential equations models in biology, epidemiology and ecology*. Nova York: Springer.
- BYBEE, J. (2010). *Language, usage and cognition*. Cambridge: Cambridge University Press.
- CASE, T. J. (1999). *An illustrated guide to theoretical ecology*. Nova York: Oxford University Press.
- CASTELLANO, C.; FORTUNATO, S.; LORETO, V. (2009). «Statistical physics of social dynamics». *Reviews of Modern Physics*, núm. 81, pàg. 591-646.
- CASTELLÓ, X.; EGUÍLUZ, V. M.; SAN MIGUEL, M. (2006). «Ordering dynamics with two non-excluding options: bilingualism in language competition». *New Journal of Physics*, núm. 8, pàg. 308.
- CASTELLÓ, X.; TOIVONEN, R.; EGUÍLUZ, V. M.; SARAMKÄI, J.; KASKI, K.; SAN MIGUEL, M. (2007). «Anomalous lifetime distributions and topological traps in ordering dynamics». *Europhysics Letters*, núm. 79, pàg. 1-6.
- CHAPEL, L.; CASTELLÓ, X.; BERNARD, C.; DEFFUANT, G.; EGUÍLUZ, V. M.; MARTIN, S.; SAN MIGUEL, M. (2010). «Viability and resilience of languages in competition». *PLoS ONE*, núm. 5, pàg. e8681.
- CURTIS, V. (2018). *Online citizen science and the widening of academia*. Cham: Palgrave Macmillan.
- DE MARCHI, S. (2005). *Computational and mathematical modeling in the social sciences*. Nova York: Cambridge University Press.
- DOREIAN, P.; HUMMON, N. P. (1976). *Modeling social processes*. Nova York: Elsevier Scientific Publishing Company, Inc.
- FARARO, T. F. (2011). «Generative process model building». A: DEMEULENAERE, P. (2011). *Analytical sociology and social mechanisms*. Cornwall: Cambridge University Press, pàg. 99-120.
- FREEDMAN, H. I. (1980). *Deterministic mathematical models in population ecology*. Nova York: Marcel Dekker, Inc.
- FUJIE, R.; AIHARA, K.; MASUDA, N. (2013). «A model of competition among more than two languages». *Journal of Statistical Physics*, núm. 151, pàg. 289-303.
- GRIMALDO, F.; SÁNCHEZ, C.; LÓPEZ-IÑESTA, E.; VICENTE, J. E.; QUEROL, E. (2017). «Ciència ciutadana contra prejudicis lingüístics: el projecte Milmots.eu». *Llengua, Societat i Comunicació: Revista de Sociolingüística de la Universitat de Barcelona*, núm. 15, pàg. 48-53. <https://revistes.ub.edu/index.php/LSC/article/view/18071>.
- GRIMALDO, F.; LÓPEZ-IÑESTA, E.; PERUCHO, M.; QUEROL, E. (2016). «Predicció de l'ús del català mitjançant la classificació supervisada». *Treballs de Sociolingüística Catalana*, núm. 26, pàg. 181-197. <http://publicacions.iec.cat/repository/pdf/00000238/00000088.pdf>.

- GRIN, F. (2003). «Language planning and economics». *Current Issues in Language Planning*, vol. 4, núm. 1, pàg. 1-66.
- HERDINA, F.; JESSNER, U. (2002). *A dynamic model of multilingualism. Perspectives of change in psycholinguistics*. Buffalo / Toronto / Sidney: Multilingual Matters Clevedon.
- IRWIN, A. (2002). *Citizen science: A study of people, expertise and sustainable development*. Abindong / Nova York: Routledge.
- JOHN, A.; YI, K.-M. (1995). «Language, learning, and location». *Thomas Jefferson Center for Political Economy Discussion Paper*, núm. 264.
- KABATEK, J. (2012). «Modelos matemáticos e substitución lingüística». *Estudos de Lingüística Galega*, núm. 4, pàg. 27-43.
- KREMER-MARIETTI, A.; DHOMBRES, J. (2006). *L'épistémologie: État des lieux et positions*. París: Ellipses.
- KRETZSCHMAR, W. (2015). *Language and complex systems*. Cornwall: Cambridge University Press.
- LÓPEZ-IÑESTA, E.; QUEIRUGA-DIOS, M. Á.; GARCÍA-COSTA, D.; GRIMALDO, F. (2021). «Citizen science projects: an opportunity for scientific literacy and sustainability education». *Metode Science Studies Journal*, núm. 12. <https://dx.doi.org/10.7203/metode.12.17824>.
- MACY, M. W.; WILLER, R. (2002). «From factors to actors: computational sociology and agent-based modeling». *Annual Review of Sociology*, vol. 28, pàg. 143-166.
- MINETT, J. W.; WANG, W. S.-Y. (2008). «Modelling endangered languages: the effects of bilingualism and social structure». *Lingua*, núm. 118, pàg. 19-45.
- MIRA, J.; PAREDES, A. (2005). «Interlinguistic similarity and language death dynamics». *Europhysics Letters*, núm. 69, pàg. 1031-1034.
- MIRA, J.; SEOANE, L. F.; NIETO, J. J. (2011). «The importance of interlinguistic similarity and stable bilingualism when two languages compete». *New Journal of Physics*, núm. 13, pàg. 033007.
- MIRALLES, C.; PERUCHO, M.; QUEROL, E. (2015). «Models dinàmics de competició entre llengües aplicats al cas català-castellà al País Valencià». *Anuari de Psicologia de la Societat Valenciana de Psicologia*, vol. 16, núm. 1, pàg. 31-55. <http://roderic.uv.es/handle/10550/53718>.
- MIRALLES, C.; PERUCHO, M.; QUEROL, E.; GRIMALDO, F.; LÓPEZ-IÑESTA, E. (2017). «La modelització matemàtica en la sociologia del llenguatge: estudi comparatiu de l'evolució dels usos lingüístics a Catalunya i al País Valencià». *Llengua, Societat i Comunicació: Revista de Sociolingüística de la Universitat de Barcelona*, núm. 15, pàg. 94-101. <https://revistes.ub.edu/index.php/LSC/article/view/18107>.
- PATRIARCA, M.; HEINSALU, E. (2009). «Influence of geography on language competition». *Physica A*, núm. 388, pàg. 174-186.
- PATRIARCA, M.; LEPPÄNEN, T. (2004). «Modeling language competition». *Physica A*, núm. 338, pàg. 296-299.

- PINASCO, J. P.; ROMANELLI, L. (2006). «Coexistence of languages is possible». *Physica A*, vol. 361, núm. 1, pàg. 355-360.
- QUEROL, E. (2000). *Els valencians i el valencià: Usos i representacions socials*. Paiporta: Denes.
- QUEROL, E. (2001). «Evolució dels usos i de les representacions socials de les llengües a Catalunya (1993-2000)». *Noves SL*, núm. 2, estiu. www6.gencat.net/llengcat/noves/hm01estiu/catalana/querolb1_10.htm.
- QUEROL, E. (2002). «Usos i representacions de les llengües a les Illes Balears». *Noves SL*, núm. 6, estiu. www6.gencat.net/llengcat/noves/hm02estiu/catalana/querol1_12.htm.
- QUEROL, F.; PELACHO, M.; CLEMENTE, J.; IBAÑEZ, M. C.; GUARDIA, L.; LISBONA, D.; SEVILLA-CALLEJO, M. (2020). *Informe del Observatorio de la Ciencia Ciudadana en España 2020*. https://ciencia-ciudadana.es/wp-content/uploads/2021/01/informe-observatorio_2020.pdf.
- SUTANTAWIBUL, C.; XIAO, P.; RICHIE, S.; FUENTES-RIVERO, D. (2018). «Revisit language modeling competition and extinction: a data-driven validation». *Journal of Applied Mathematics and Physics*, vol. 6, núm. 7.
- TEMPLIN, T. (2019). «A language competition model for new minorities». *Rationality and Society*, vol. 31, núm. 1, pàg. 40-69.
- TOIVONEN, R.; CASTELLÓ, X.; EGUÍLUZ, V. M.; SARAMÄKI, J.; KASKI, K.; SAN MIGUEL, M. (2009). «Broad lifetime distributions for ordering dynamics in complex networks». *Physics: Review E*, núm. 79, pàg. 016109.
- VERSPOOR, M.; DE BOT, K.; LOWIE, W. (2011). *A dynamic approach to second language development: Methods and techniques*. Amsterdam: John Benjamins Publishing.
- WANG, W. S.-Y.; MINETT, J. W. (2005). «The invasion of language: emergence, change and death». *Trends in Ecology and Evolution*, núm. 20, vol. 5, pàg. 263-269.
- WILDGEN, W. (1982). *Catastrophe theoretic semantics: An elaboration and application of René Thom's theory*. Amsterdam: John Benjamins Publishing.
- WYBURN, J.; HAYWARD, J. (2008). «The future of bilingualism: an application of the Baggs and Freedman model». *Journal of Mathematical Sociology*, núm. 32, pàg. 267-284.
- ZHANG, M.; GONG, T. (2013). «Principles of parametric estimation in language competition». *Proceedings of the National Academy of Sciences*, vol. 110, núm. 24, pàg. 9698-9703.

Canvi lingüístic i anivellament dialectal¹

Maria-Pilar Perea
Universitat de Barcelona

1. INTRODUCCIÓ

El títol d'aquest article és volgudament ampli i reflecteix, en certa manera, una obvietat. El canvi lingüístic és una manifestació més d'aquesta transformació que s'esdevé en qualsevol ésser, estat, fenomen o situació i que es troba condicionat principalment pel factor temps. I atès que la realitat canvia de manera inexorable, fora bo que es poguessin determinar les causes o els principis que generen els canvis; i, és més, fins i tot predir les transformacions que poden tenir lloc en el futur i els resultats que aquestes poden provocar. I ja no dic si fora possible aturar, en la mesura que sigui possible, certs processos no desitjats de canvi que estan en marxa.

Des de l'àmbit de la lingüística, el canvi s'ha definit com la variació que poden experimentar els diversos aspectes i trets que componen la llengua, els quals, per a una comprensió més gran i per a un estudi més acurat i sistemàtic, s'han distribuït en els diferents compartiments en què s'ha dividit aquesta disciplina. Així, es parla de canvis de tipus fonètic, morfològic, sintàctic o lèxic, per bé que poden estar interrelacionats. Aquests canvis, a més del pas del temps i de la diferenciació motivada per la situació geogràfica, poden ser provocats o afavorits per fenòmens —endògens o exògens, lingüístics o extralingüístics— com ara l'economia (o la llei del mínim esforç), l'analogia, el contacte de llengües, l'ensenyament, els mitjans de comunicació, etc. I també, naturalment, per factors psicolingüístics, com ara el prestigi, l'autoodi, la retracció, etc.

Aquest treball se centrarà particularment en la morfologia verbal regular de tres varietats dialectals del català: el català occidental (o nord-occidental), el valencià i el balear. L'objectiu és doble: en primer lloc, donar compte dels processos de canvi que han afectat els diferents morfs (i de vegades les arrels)

¹ Aquest treball s'adscriu al projecte PGC2018-095077-B-C43, finançat pel Ministeri de Ciència, Innovació i Universitats.

que componen els verbs que paradigmàticament s'han considerat models regulars; i, en segon lloc, descriure, d'una banda, la naturalesa del canvi, que, com és previsible, pot tendir o bé cap a l'anivellament dialectal (és a dir, i en altres mots, cap a l'adopció del model de la varietat estàndard, cosa que implica una desdialectalització) o bé cap a un increment de les diferències dialectals (cosa que implicaria una dialectalització); i de l'altra, determinar, si és possible, les motivacions del canvi.

Els verbs que s'han considerat regulars són, seguint la divisió proposada per Viaplana (1986) i que ja es va recollir en la *Gramàtica del català contemporani* (2008: 597 i seg.), per bé que amb lletra petita:

- *cantar*, com a model de la primera (I) conjugació;
- *perdre* i *beure*, com a models de la segona (II) conjugació, sense extensió i amb extensió, respectivament, i
- i *sentir* i *servir* (*partir*), que corresponen al model de la tercera (III) conjugació, amb la divisòria, ben coneguda, entre verbs no incoatius i verbs incoatius.

L'elecció de *beure*, en general proscrit a la irregularitat per la seva extensió velar (en la *Gramàtica de la llengua catalana* [GIEC] i en la *Gramàtica bàsica i d'ús de la llengua catalana* [GBU] rep aquesta consideració, veg. p. 252-254 i p. 101, respectivament), respon al fet, ja molts cops argumentat, que tot i que es tracta d'una subconjugació amb pocs representants (malgrat tot, numèricament són més que els verbs que componen la subconjugació erròniament dita pura o no incoativa), els morfs corresponents a l'extensió velar, que caracteritzen aquesta subclasse flexiva, obviant-ne, naturalment, els orígens històrics, es troben distribuïts de manera sistemàtica en l'estructura formal de tots els verbs que segueixen aquest (sub)model.

Cal recordar que les gramàtiques normatives de l'Institut d'Estudis Catalans (les ja anomenades GIEC i GBU) consideren que els models regulars són *cantar* (I), *perdre* o *témer* (un verb, en realitat, que costa de veure com a model, atès l'ús tan restringit que se'n fa) (II), seguint, naturalment, la proposta fabriana, i *dormir* i *servir* (III). *Beure* és descrit en l'apartat «Verbs amb alternances en el segment final del radical». Tanmateix, si en parlar de conjugacions es consulta la GBU, més simplificada, s'explicita que «[la] conjugació es manifesta, generalment, a través de la vocal temàtica» (p. 82). Les vocals temàtiques de la segona conjugació són *a*, *e* i *u*; és a dir, les que apareixen tant a *batre* com a *beure*; per tant, és indubtable que *beure* pertany a aquesta conjugació, però inclou un segment velar, que en fa una subcategoria.

Metodològicament, com a material d'anàlisi s'han utilitzat els materials de dos aplecs de dades de morfologia verbal, els quals poden ser contrastats cronològicament, atès que es van recopilar amb una distància temporal d'aproximament una vuitantena llarga d'anys; és a dir, que si el període d'una generació es considera que abraça un interval entre vint-i-cinc i trenta anys, hi ha una diferència, doncs, entre les dades, de tres generacions. D'una banda, em refereixo al tresor que ofereix *La flexió verbal en els dialectes catalans (LaFV)*, obtingut, entre el 1906 i el 1926, per Antoni M. Alcover i que fou sistematitzat i editat per Francesc de B. Moll; i, de l'altra, les dades del Corpus Oral Dialectal (COD), que es van recollir en la dècada dels anys noranta del segle passat com a resultat d'un projecte desenvolupat en el Departament de Filologia Catalana de la Universitat de Barcelona.

Cal admetre que la hipòtesi de partida considerava que les dades (relativament) recents, les recollides al COD, encara que només fos per raons temporals havien de discrepar d'una manera significativa de les dades alcoverianes i havien de tendir cap a les realitzacions de l'estàndard. No hi havia cap hipòtesi pel que fa a quina varietat, de les tres examinades, se n'allunyava més, però se suposava que devien ser el valencià o el balear.

Per bé que es podria oferir una visió panoràmica de les tendències de canvi en la morfologia verbal dialectal en totes les varietats de català, la selecció de les tres varietats ha respost a dues motivacions pragmàtiques: la primera, òbviament, als límits que ha de tenir aquest treball; la segona, al fet que aquestes varietats, especialment les dues darreres, han rebut més atenció per part de les propostes normatives, les quals permeten —toleren, segons com— algunes de les seves peculiaritats i sancionen determinats comportaments que han de ser substituïts naturalment pel model de l'estàndard. Vegeu, per exemple, el símbol de la maneta, pel que fa a les remarques normatives i d'ús, que apareix en la *Gramàtica bàsica i d'ús de la llengua catalana* de l'IEC (p. 73-83) i que acostuma a referir-s'hi. Les al·lusions a les solucions occidentals són molt menors, i d'aquí es deriva el prejudici que he esmentat més amunt. Veurem si s'acompleixen les presumpcions indicades.

2. CORRELACIÓ ENTRE CANVI LINGÜÍSTIC I ANIVELLAMENT DIALECTAL

Abans de començar la descripció i la comparació d'alguns aspectes de la morfologia verbal que han experimentat modificacions des d'un punt de vista formal i estructural, cal fer una breu referència al concepte de canvi lingüístic considerat des de la perspectiva de l'anivellament dialectal.

Ja s'ha comentat, i no cal remuntar-se a concepcions de filòsofs com Aristòtil o Heràclit, que el canvi és connatural a qualsevol entitat, situació o fenomen; i hi ha qui s'ha preguntat (vegeu Aitchinson, 2013) si, aplicat a les llengües, el canvi implica un procés progressiu vers la decadència o la degeneració («Tongues, like governments, have a natural tendency to degeneration», Samuel Johnson, 1755, *ap.* Aitchinson, 2013: 9), si es parteix d'una visió purista; o, per contra, si implica un progrés d'enriquiment que les faria més eficients («In the evolution of languages the discarding of old flexions goes hand in hand with the development of simpler and more regular expedients that are rather less liable than the old ones to produce misunderstanding», Otto Jespersen, 1922, *ap.* Aitchinson, 2013: 7). Caldria considerar, però, quin significat i quin abast tindria aquesta presumpta eficiència.

Un bon nombre de treballs s'han desenvolupat fora de les nostres fronteres en relació amb les conseqüències de l'anivellament dialectal (*dialect levelling*), concepte que també s'ha equiparat amb el resultat de fenòmens de convergència lingüística (Heeringa i Nerbonne, 2001) pel que fa a la pressió que pot exercir l'estàndard en una determinada varietat. Vegeu, entre molts d'altres, els estudis de Kerswill (2003) i Britain (2010). En un treball relativament recent, Auer (2018: 159) declarava: «There can be no doubt that the main development in the European dialects over the past hundred years or so has been convergence, both among the dialects themselves and toward the standard variety».

Pel que fa al català, diversos estudis han fet referència als processos de canvi lingüístic que han donat lloc a un anivellament dialectal, o que es troben en procés d'aconseguir-ho. Aquests treballs han incidit no sols en varietats o àrees no estudiades en aquest treball (per exemple, vegeu, per al mallorquí, Radatz, 2007; per al salat, Perea, 2007a; per al rossellonès, Perea, 2007b i 2009a; per a l'alguerès, Perea, 2010, o per al calero, Lladó i Borràs, 2012), sinó també, i especialment, en relació amb el català occidental (vegeu, per exemple, els treballs de Massanell, 2001 i 2012, on s'utilitza el concepte de procés d'orientalització —és obvi que el català oriental, o central, d'acord amb les diverses etiquetes terminològiques, és la base de l'estàndard—, que desemboca en l'estandardització; i també diversos estudis de Valls, tant pel que fa a la zona esmentada [Valls, 2008 i 2013] com els referits a la franja aragonesa [Valls, 2017 i 2019], i amb l'aplicació de sistemes quantitius d'anàlisi). També és interessant, des d'una perspectiva general, el treball de Kabatek i Pusch (2009).

3. LES DADES DE REFERÈNCIA, LES ÀREES DIALECTALS I LES LOCALITATS

A continuació, es farà una presentació molt breu dels materials objecte de comparació.

La flexió verbal en els dialectes catalans és un corpus de gairebé 500.000 formes verbals que recull la conjugació completa de 70 verbs en 149 localitats de tot el domini lingüístic català. Els verbs foren objecte de l'enquesta feta per Antoni M. Alcover entre el 1906 i el 1928. Francesc de B. Moll publicà el treball entre el 1929 i el 1932. Els materials han estat informatitzats, cartografiats i completats a Perea (1999, 2001 i 2005a) i ara es poden consultar al Portal Alcover de l'Institut d'Estudis Catalans (alcover.iec.cat). Aquest és el recull més complet que el català té pel que fa a la morfologia verbal dialectal de començament del segle xx.

El COD conté dades fonètiques i morfològiques obtingudes mitjançant qüestionaris i recull també un conjunt d'enregistraments de parla espontània d'uns deu minuts de durada. Les enquestes es van fer entre el 1995 i el 1997 en 106 capitals de comarca o localitats equivalents de tots els territoris de parla catalana. Pel que fa a la morfologia verbal, es va enquestar la conjugació completa de cinc verbs que es consideraven paradigmàtics de la regularitat verbal: *cantar* (I), *perdre* (II, -ext), *beure* (II, +ext), *sentir* (III, -ext) i *servir* (III, +ext).²

Les tres àrees dialectals, per la seva naturalesa, presenten variació lingüística, que és motivada principalment per la seva ubicació en zones geogràfiques diferents. A aquest fet, cal afegir-hi l'existència de determinades variables socials, que van ser diferents en cadascun dels projectes³ però que es troben vinculades per un objectiu comú: la pretensió que les dades fossin representatives per tal que donessin compte de la realitat lingüística del moment. Diafàsicament, doncs, les realitzacions responien no pas al codi escrit, sinó a l'oral vulgar, en el sentit de general, informal o col·loquial.

Per dur a terme l'estudi s'han seleccionat les localitats comunes als dos aplecs de dades, que poden ser susceptibles de comparació. En voler fer una anàlisi diacrònica de materials de projectes que tenen una semblança formal —per exemple, atles o conjunts de dades—, el que desitjaria l'investigador és que les localitats estudiades i la tipologia dels materials enquestats coincidissin. Si és així, es pot fer un contrast i una comparació magnífics, però aquest

² El verb *servir* no és a *LaFV*; per aquest motiu s'ha pres com a referència el verb *partir*, que també encaixa en la classe (III, + ext).

³ Vegeu les característiques dels informadors de *La flexió verbal* a Perea (2005b) i les dels del COD a Lloret i Perea (2002).

fet, per diversos motius, no acostuma a succeir. Penso, per exemple, en les localitats de l'*Atlas Lingüístico de la Península Ibérica* (ALPI) i en les de l'*Atlas Lingüístico del Domini Català* (ALDC), entre les quals, per voluntat dels qui van concebre aquest darrer projecte, hi va haver una coincidència mínima. Per aquest motiu, s'han seleccionat, de les tres àrees dialectals objecte d'estudi, les localitats següents, que són les que coincideixen en tots dos projectes:

- a) Català occidental (14 localitats):⁴ (66) Andorra; Aragó: (77) Benavarri, (81) Tamarit, (88) Fraga; Lleida: (71) el Pont de Suert, (72) Sort, (73) la Seu d'Urgell, (76) Tremp, (82) Balaguer, (84) Cervera, (86) Lleida; Tarragona: (91) Falset, (92) Gandesa, (94) Tortosa.
- b) Balear (9 localitats): Mallorca: (118) Pollença, (121) Sóller, (129) Palma, (130) Manacor, (131) Felanitx; Menorca: (133) Ciutadella, (138) Maó; Eivissa: (146) Eivissa; Formentera: (147) Sant Francesc Xavier de Formentera.
- c) Valencià (13 localitats): Castelló: (95) Morella, (96) Vinaròs, (100) Castelló; València: (101) Sagunt, (102) Llíria, (103) València, (104) Sueca, (105) Gandia; Alacant: (109) Cocentaina, (112) Alcoi, (116) Alacant, (114) la Vila Joiosa, (117) Elx.

4. LES TENDÈNCIES DEL CANVI LINGÜÍSTIC EN LA MORFOLOGIA VERBAL REGULAR

En primer lloc, per determinar quina tendència ha experimentat la morfologia verbal regular en les diverses localitats examinades, es farà una comparació contrastiva de les dades. Els temps i modes verbals seran el punt de partida, fet que permetrà, secundàriament, considerar quin són els més proclius a mostrar variació, i amb quina recurrència, i si hi ha temps i modes més estables; i, alhora, quins són més susceptibles a l'estandardització o a la dialectalització.

S'exemplificaran sols els casos més significatius i representatius, ja que l'aplec de dades resultant és realment molt considerable. Si algú suposa que hi pot haver una certa coincidència entre els resultats de les localitats d'una mateixa àrea, s'equivoca del tot. Sí que es pot determinar, però, a la vista dels resultats, si hi ha àrees, i sobretot localitats, que tenen una morfologia verbal regular més estandarditzada i que presenten, per tant, menys variació morfològica de caràcter divergent en relació amb la proposta normativa.

⁴ Es fa servir la numeració que s'aplica en *LaFV*.

4.1. Aspectes preliminars: les tendències fonètiques

No es pot estudiar tampoc, perquè superen l'abast d'aquest estudi i, per tant, solament es poden apuntar, algunes qüestions fonètiques que s'han evidenciat en fer la comparació dels materials. Es tracta, en relació amb el vocalisme, de la realització (o la transcripció) del timbre d'algunes vocals finals, que és molt més detallada en *LaFV* que en COD, cosa que en certa manera podria fer pensar en una simplificació vocàlica. Tampoc no hi ha coincidència en la realització de la *a* tònica a Felanitx (documentada, entre altres autors, a Grimalt [1972] i a Bibiloni [2016]), de caràcter més tancat i anterior, que *LaFV* transcriu [æ], mentre que el COD, [á]. I així es constata que hi ha variació en formes com [cəntæw] – [cəntáw].

Pel que fa al consonantisme, l'acarament de les dades dona compte de processos de canvi i de variació ben interessants; per exemple, la pèrdua de la labiodental en determinades localitats valencianes o occidentals. Observeu les alternances que mostrava *LaFV* a Falset, [beβén ~ bevén], i que el COD unifica en [beβén]. O la substitució de la labiodental a Castelló: *LaFV* [bevén] – COD [beβén]; a Elx: *LaFV* [bevént] – COD [beβént]; a Tortosa: *LaFV* [kantáva] – COD [kantáβa]; a Vinaròs: *LaFV* [kantáve] – COD [kantáβa].

En relació amb l'allunyament de l'estàndard pel que fa a aspectes morfològics, no es consideren rellevants trets que fonèticament es poden explicar, en la llengua oral, per processos d'assimilació o de dissimilació i que no afecten l'estructura morfològica de les formes verbals. Les realitzacions fonètiques responen a processos fonològics que incideixen sobre formes bàsiques o abstractes comunes. Aquests processos solen ser de caràcter general, però també podrien ser de caràcter individual, en relació amb l'informador. Vegeu, per exemple, a Sóller, l'assimilació que es produeix en el participi del verb *beure*, *LaFV* [buyút] – COD [bəyút], la qual, com es veu, tendeix en darrer terme a la solució de l'estàndard. El mateix procés també es troba en la segona persona del singular de l'imperatiu a Felanitx: *LaFV* [bów] – COD [bów] ~ [bów]. També a Sóller es detecta una labialització i una assimilació: *LaFV* [bówro, ~ bówrə] – COD [báwrə ~ bówrə].

En relació amb el comportament de la ròtica, l'infinitiu del verb *perdre* també pot experimentar certes dissimilacions:

el Pont de Suert	<i>LaFV</i> [pérde]	COD [pérðe ~ pérðre]
Balaguer	<i>LaFV</i> [pérdre ~ pérdiri]	COD [pérðe ~ pérðre]
Gandesa	<i>LaFV</i> [pérdre]	COD [péðre ~ pérðre]
Alacant, Alcoi,	<i>LaFV</i> [pédre]	COD [péðre ~ pérðre]
Elx, Vinaròs		

El fenomen de dissimilació, com també l'addició d'una vocal epentètica per reparar la seqüència d'una coda i d'una obertura complexes, es pot trobar també en el futur i el condicional del verb *perdre*:

Localitat	LaFV	COD
Benavarri Fraga Lleida	perdré perdrás perdrá	perðeré ~ perðré perðerás ~ perðrás perðerá ~ perðrá
Elx	pedré pedrás	peðré ~ perðré
Morella	pedré ~ perderé pedrás ~ perderás	perðeré perðerás
Sóller	pæddré	pæðæré ~ pæðré
Manacor	bæwré	bæwré ~ bowré
Sóller	bæwré bæwrás	bæwré bowré
Vinaròs	pedríe pedríes pedríe	peðría ~ perðría peðríes ~ perðríes peðría ~ perðría
Morella	pedría ~ perdería...	perðería...

Finalment, es constata també una assimilació en el radical del verb *sentir*:⁵

Localitat	LaFV	COD
Benavarri	sentiré sentirás sentirá	sentiré ~ sintiré sentirás ~ sintirás sentirá ~ sintirá

⁵ Aquesta assimilació també es pot trobar en altres temps i en altres localitats: en les dades de LaFV a Balaguer [sentíw ~ sintíw], però no en les del COD.

4.2. Les tendències del canvi

En examinar comparativament les dades dels dos corpus s'han observat dues tendències generals que coincideixen, en un grau major o menor, en tots els dialectes examinats:

- a) l'estandardització, que pot implicar, però no necessàriament, una simplificació paradigmàtica;
- b) la divergència (dialectalització), que pot comportar una diversificació paradigmàtica.

A continuació, es presenten les dades de manera conjunta, ja que en l'apartat següent es quantificarà, en funció de cada varietat dialectal, el grau de convergència i de divergència en relació amb les solucions de l'estàndard.

Els casos més simples els constitueixen dues formes impersonals que mostren fenòmens prou sistemàtics quant a les dues tendències esmentades.

D'una banda, pel que fa a l'infinitiu, s'aprecia:

- a) La reducció de formes alternatives a favor de la realització estàndard: s'observa la presència de la vocal *i* final en el cas del verb *perdre* i la labialització en el radical del verb *beure*.
- b) L'alternança de solucions, motivada per l'adjunció d'una marca d'infinitiu doble, que suposa una «hipercharacterització», seguint la terminologia de Malkiel (1957-1958: 101).

Gandia *LaFV* [pédre] / [béwre] COD [péðre ~ péðrer] / [béwre ~ béwrer]

En el participi s'aprecien també les dues tendències indicades:

- a) L'estandardització, definida en aquest cas per l'absència o la presència de la velarització:

Llíria, Morella,	<i>LaFV</i> [perdút ~ perýút]	COD [perðút]
Sagunt, Sueca, València		
Cervera	<i>LaFV</i> [beyút ~ beút]	COD [beyút]

- b) La difusió de la velarització, mitjançant un fenomen que es repetirà al llarg de la descripció contrastiva dels diferents temps:

Alcoi	<i>LaFV</i> [perdút]	COD [perdút ~ perýút]
-------	----------------------	-----------------------

El condicional presenta molt poca variació tant en les dades de *LaFV* com en les del COD. Sols mostra, en aquest darrer aplec de dades, l'adopció de les formes estàndard mitjançant la generalització de la desinència de mode/temps, que es trobava simplificada en alguns indrets del català occidental:

Localitat	<i>LaFV</i>	COD
la Seu d'Urgell ⁶	bewríə bewrís bewrí bewrím bewríu bewrín	bewría ~ bewríə bewríes bewríe bewríem bewríew bewríen
Tremp	bewríe bewríes ~ bewrís bewrí ~ bewríe bewrím ~ bewríu bewríen ~ bewrín bewríen	bewría bewríes bewría ~ bewríe bewríem bewríew bewríen

A continuació es descriuen els temps següents, que mostren paradigmes complexos susceptibles d'experimentar una variació considerable.

4.2.1. Gerundi

a) Tendència a l'estandardització

En alguns indrets del català occidental, s'observa una reducció del polimorfisme i una tendència a fer servir les realitzacions de l'estàndard.

Cervera	<i>LaFV</i> [beβén ~ beyén ~ bején]	COD [beβén]
Lleida	<i>LaFV</i> [beβén ~ bején]	COD [beβén]

Aquesta tendència pot implicar, al seu torn, que les formes dialectals convisquin amb les formes estàndard, fins i tot en casos en què en les dades de *LaFV* es partia únicament d'aquestes darreres realitzacions:

⁶ Aquest paradigma, llevat del timbre de la vocal de la primera persona, es reproduïx, segons les dades de *LaFV*, també a Sort, Tremp i Andorra.

Sort	<i>LaFV</i> [bején]	COD [beén ~ beyén]
València	<i>LaFV</i> [beþént]	COD [beþént ~ beyént]
Gandia	<i>LaFV</i> [perdént]	COD [perðént ~ peryént]
Pollença	<i>LaFV</i> [pærdént]	COD [pærðént ~ pærðijént]
Manacor	<i>LaFV</i> [səntínt]	COD [səntínt ~ səntijént]
Maó i Sóller	<i>LaFV</i> [pærtínt]	COD [særvínt ~ særvijént]

b) Tendència a la divergència

En alguns indrets es parteix, en general, de respostes alternatives, una de les quals fins i tot pot ser l'estàndard. El COD recull formes velaritzades, amb caiguda d'aquest element aproximant o amb l'adjunció d'una semivocal antihiatíca. Observem en alguns casos la difusió de la velarització.

la Seu d'Urgell	<i>LaFV</i> [beþén ~ beyén]	COD [beyén]
Balaguer	<i>LaFV</i> [beþén ~ bején]	COD [bején]
Alcoi	<i>LaFV</i> [bevént]	COD [beén ~ beyén]
Ciutadella	<i>LaFV</i> [bəvént]	COD [bəvént]
Maó	<i>LaFV</i> [bəént ~ bævént ~ bævént]	COD [bəvént]

4.2.2. Imperatiu

Moll dona compte de la relació que existeix, des d'un punt de vista històric, entre el subjuntiu, l'indicatiu i l'imperatiu: el «llatí vulgar va substituir les primeres i terceres persones de l'imperatiu per les respectives del present de subjuntiu, i la segona del plural per la del present d'indicatiu» (Moll, 1991, *ap.* Grandgent, LV, § 412).

En les dades de l'imperatiu estudiades s'observa una tendència a l'estandardització:

Castelló	<i>LaFV</i> [sén ~ sentís]	COD [sén]
----------	----------------------------	-----------

- Convivència entre formes dialectals i estàndard:

Sort	<i>LaFV</i> [sín]	COD [sén ~ sentí] ~ sín]
Vinaròs	<i>LaFV</i> [partí]	COD [serþís ~ serþí]

• En la segona persona del singular, s'observa un canvi en l'increment dels verbs de la tercera conjugació:

Lleida, Andorra, Cervera	<i>LaFV</i> [partíj]	COD [serβéj]
Tremp	<i>LaFV</i> [partís]	COD [serβéj]

Les mateixes tendències s'aprecien en les primeres i segones persones de l'imperatiu:

a) Tendència a l'estandardització (en la minoria de casos)

Cervera	<i>LaFV</i> [beɣém ~ bejém]	COD [beβém]
Andorra	<i>LaFV</i> [partiyém]	COD [serβím]

• Convivència de formes alternatives

Tortosa	<i>LaFV</i> [perdém/w]	COD [perðém/w ~ perɣém/w ~ perɣésem/w]
Balaguer	<i>LaFV</i> [sentím/w ~ sintím/w]	COD [sentím/w ~ sentiyém/w]
Benavarri	<i>LaFV</i> [perdám/w]	COD [perðám/w ~ perðém/w]
Pollença, Felanitx, Manacor	<i>LaFV</i> [pərdám/w]	COD [pərdám/w ~ pərdijém/w]
Pollença, Manacor	<i>LaFV</i> [pərtím/w]	COD [sərvím/w ~ sərvijém/w]
Palma	<i>LaFV</i> [pərtím/w]	COD [sərvijém/w ~ sərvijém/w]

b) Tendència a la divergència per raó de la difusió de la velarització. Cal observar que algunes de les formes recollides en *LaFV* corresponien a les solucions estàndard.

Benavarri	<i>LaFV</i> [beɣwém/w]	COD [beɣám ~ beɣwém/w]
Alcoi, Gandia	<i>LaFV</i> [perðém/w]	COD [perðém/w ~ perɣám/w ~ perɣém/w]
Vinaròs	<i>LaFV</i> [perdém/w]	COD [perɣém/w ~ perɣésem/w]
Vinaròs	<i>LaFV</i> [partiyém]	COD [serβím ~ serβiyém ~ serβiskésem]
Elx, Sagunt, València	<i>LaFV</i> [sentím/w]	COD [sentím/w ~ sintám/w]
Alacant, Castelló	<i>LaFV</i> [perdém/w]	COD [perɣém/w]
Eivissa, Ciutadella, Sóller	<i>LaFV</i> [bəóm]	COD [bəɣóm]
Felanitx	<i>LaFV</i> [pərtím/w]	COD [sərvijém/w]

S'observa una vacil·lació en el timbre de les vocals temàtiques tant de la primera com de la segona conjugacions: Felanitx: *LaFV* [cəntám] - COD [cəntám ~ cəntém].

Cal tenir en compte que, en general i llevat d'alguna excepció, es pot establir aquesta relació entre el present de subjuntiu i l'imperatiu, exemplificada amb el verb *cantar* en balear:

Imperatiu	Present de subjuntiu	Present d'indicatiu
	canti	cant
canta (tu)	cantis	cantes
canti (ell)	canti	canta
cant[ó]m	cant[ó]m	cant[á]m
cant[á]w	cant[ó]w	cant[á]w
cantin	canti	canten

La mateixa vacil·lació apareix en alguns indrets quant a l'aparició del segment velar, que, d'acord amb la relació anterior, hauria d'aparèixer en la primera persona del singular de l'imperatiu, atès que prové del subjuntiu:

Castelló: *LaFV* [bevém ~ bevév] - COD [beyém ~ beyév].

4.2.3. Present d'indicatiu

a) Tendència a l'estandardització

- Canvi d'extensió (ja indicada en la segona persona del singular a § 3.2.2)

Localitat	<i>LaFV</i>	COD
Andorra	partíʃo	serβéjʃo
Cervera	partíʃes	serβéjʃes
Lleida	partíʃ	serβéjʃ

- Convivència de solucions

Localitat	LaFV	COD
el Pont de Suert Tremp	partíso partíses partís partísen	serβéjfo ~ serβíso serβéjjes ~ serβíses serβéjɿ ~ serβís serβéjjen ~ serβísen
Vinaròs	partíɿk ~ partífo partíjes partíɿ	serβísko ~ serβíso ~ serβífo serβíses ~ serβíjes ~ serβískes serβís ~ serβíɿ

En balear té lloc un canvi de desinència:

Localitat	LaFV	COD
Felanitx, Manacor, Sóller	pərtájs	sərváɿəs
Palma, Pollença	pərtájs ~ pərtáɿəs	serváɿəs

Les formes de *LaFV* eren admissibles d'acord amb la *Proposta per a un estàndard oral*.

b) Tendència a la divergència, amb l'aparició de solucions polimòrfiques.

Localitat	LaFV	COD
Alacant	séɿk séns...	séɿk ~ sentísk séns ~ sentíɿes...
Vinaròs	partíɿk ~ partífo partíjes partíɿ...	serβísko ~ serβíso ~ serβífo serβíses ~ serβíjes ~ serβískes serβís ~ serβíɿ...

- Pèrdua de l'extensió

Localitat	LaFV	COD
Castelló	séɿk ~ sentísk séns ~ sentíses sén ~ sentís...	séɿk séns sén...

- Adquisició d'extensió

Localitat	LaFV	COD
Fraga	sínto síns sín sínten	sentísko sentísés sentís sentísen

En la primera i la segona persones del plural del present d'indicatiu s'observen les mateixes tendències que en les mateixes persones de l'imperatiu:

a) Tendència a l'estandardització

Cervera, Lleida	LaFV [beβém/w ~ bejém/w]	COD [beβém/w]
Falset	LaFV [beβém/w ~ bevém/w]	COD [beβém/w]
Andorra	LaFV [partiým/w]	COD [serβím/w]
Morella	LaFV [beβém/w ~ bevém/w]	COD [beβém/w]
Morella	LaFV [partím/w ~ partiým/w ~ parti{kém/w]	COD [serβím/w]
Sagunt	LaFV [perðém/w ~ peryém/w]	COD [perðém/w]

- Amb convivència de formes dialectals i estàndard:

Tremp	LaFV [bejém/w]	COD [beβém/w ~ bejém/w]
Tremp	LaFV [bejém/w]	COD [beβém/w ~ bejém/w]
Alacant	LaFV [kantésem/w]	COD [kantém/w ~ kantésem/w]

En balear s'aprecia un canvi de desinència (amb labialització en alguns indrets).⁷ Les primeres labialitzacions són considerades, en la *Proposta per a un estàndard oral*, d'àmbit restringit.

Maó	LaFV [béjm] [béjs]	COD [bəvé[m] [bəvéw]
Manacor	LaFV [báj[m] ~ bəvám] [bájs ~ bəvów]	COD [bəvám ~ bovám] [bəvów ~ bovów]
Felanitx	LaFV [béjm] [béjs]	COD [bəvé[m] [bəvéw]

⁷ La labialització també es pot trobar en altres contextos, com ara el condicional a Manacor i Sóller, en les realitzacions del COD: [bəwríə] ~ [bowríə], [bəwríəs] ~ [bowríəs], [bəwríə] ~ [bowríə]. I [bəwré] ~ [bowré]... I en el present d'indicatiu a Felanitx: tu [báws] ~ [bóws], [bów] ~ [bów].

b) Tendència a la divergència, en alguns casos amb velarització

Balaguer	<i>LaFV</i> [beβém/w ~ bejém/w]	COD [bejém/w]
la Seu d'Urgell	<i>LaFV</i> [beβém/w]	COD [beyém/w]
Sueca	<i>LaFV</i> [beyém/w ~ bevém/w]	COD [beyém/w]

4.2.4. Present de subjuntiu

a) Tendència a l'estandardització

S'observa una generalització de la desinència *-i* en occidental i sobretot en balear.

Localitat	<i>LaFV</i>	COD
Andorra	béγə béγes ~ béγos béye ~ béyo beyém beyéw béγen ~ béγon	béyi béγis béyi beβém ~ beyém beβéw ~ beyéw béyin
Tremp	kánte kántes kánte kánten	kánti kántis kánti kántin
Pollença	cánt ~ cánti ~ cántiγə	cánti
Sóller	báyə ~ báyi ~ báγō báyəs ~ báγis báyə ~ báyi ~ báγō bəám bəáw báyən ~ báyin	báyí báyis báyí bəγám bəγáw báyin
Felanitx	cánt ~ cántiγə cántiγəs...	cánti cántis...
Eivissa	pérðə ~ pérði ~ pérðiγə ~ pérðiγi ~ pérγə ~ pérγi...	pérði...
Sóller	sénti ~ séntiγo, ~ séntō séntəs ~ séntiγəs ~ séntis séntə ~ sénti ~ séntiγō	sénti séntis sénti

Palma	sénti ~ séntijə séntijəs ~ séntis...	sénti séntis...
Manacor, Palma Pollença	pərtəscə ~ pərtəski...	sərvəski...

- Convivència de formes dialectals amb l'estàndard

Localitat	LaFV	COD
el Pont de Suert	kánteɣa, ~ kánteɣɔ kánteɣes ~ kántes kánte ~ kánteɣe kánteɣen ~ kánten	kánteɣa ~ kánti kánteɣes ~ kántis kánteɣe ~ kánti kánteɣes ~ kántis

- Canvis en l'arrel per creuament amb el radical de *veure*.

Localitat	LaFV	COD
Gandesa	béɣe béɣes ~ béɣos béɣe ~ béɣo beyéseɰ beyésew béɣen ~ béɣon	béɣa ~ béɣɔ béɣes ~ béɣɔs béɣe ~ béɣɔ beyéseɰ ~ beyéseɰ beyésew ~ beyésew béɣen ~ béɣon

- Canvi d'increment

Localitat	LaFV	COD
Cervera	partíʃa partíʃi partíʃis partíʃi	serβéjʃi serβíʃi serβéjʃis serβíʃis

Les dues tendències generals s'aprecien en la primera i la segona persones del plural:

a) Tendència a l'estandardització

Sueca	LaFV [seɲkám/w ~ sentím/w ~ sentám/w ~ sintam/w]	COD [sentím/w]
Morella	LaFV [sentiým/w]	COD [sentím/w]

- Convivència de solucions

Ciutadella	<i>LaFV</i> [bə́ém/w]	COD [bəvém/w ~ bə́ým/w]
Manacor	<i>LaFV</i> [pərdém/w]	COD [pərdém/w ~ pərdíým/w]
Maó	<i>LaFV</i> [pərdém/w]	COD [pərdém ~ pərdəým/w]

b) Tendència a la divergència

Sóller, Eivissa	<i>LaFV</i> [bə́ém/w]	COD [bə́ým/w]
-----------------	-----------------------	---------------

- Reducció de formes alternatives

Localitat	<i>LaFV</i>	COD
Tamarit	sínta ~ sentísko síntes ~ sentískes sínte ~ sentíske sentiým sentiýéw sínтен ~ sentísken	sínta síntes sínte sentiým sentiýéw sentísken ~ sínтен
Castelló	sénke ~ sentíske sénkes ~ sentískes sénke ~ sentíske	sénke sénkes sénke

- Canvi en les característiques de l'extensió o adopció d'un altre increment

Localitat	<i>LaFV</i>	COD
Sagunt	partíjka partíjkə partíjkes	serbíjka ~ serbíska serbíjkes ~ serbískes serbíjka ~ serbíska
la Vila Joiosa	sénggə ~ séntə séngges ~ séntes	sénga ~ sénta ~ sentíjka séngges ~ séntes ~ sentíjkes

4.2.5. Imperfet d'indicatiu

a) Tendència a l'estandardització

Localitat	<i>LaFV</i>	COD
Balaguer	kantáβe ~ kantáje kantáβes ~ kantájes kantáβe ~ kantáje kantáβem ~ kantájem kantáβew ~ kantájew kantáβen ~ kantájen	kantáβa ~ kantáβe kantáβes kantáβe kantáβem kantáβew kantáβen
Tamarit	beíβa ~ beíβo beíβes beíβe beíβem beíβew beíβen	beβía beβíes beβíe beβíem beβíew beβíen
Cervera	béja ~ beβía béjes ~ beβíes béje ~ beβíe béjam ~ beβíam béjaw ~ beβíaw béjen ~ beβíen	beβía ~ beβíə beβíes beβíe beβíem beβíew beβíen
la Seu d'Urgell	beβíβə beβíβes beβíβe beβíβom beβíβow beβíβen	beβía beβíes beβíe beβíem beβíew beβíen
Lleida	kantáβe ~ kantáje kantáβes ~ kantájes kantáβe ~ kantáje kantáβem ~ kantájem kantáβew ~ kantájew kantáβen ~ kantájen	kantáβe kantáβes kantáβe kantáβem kantáβem kantáβen
Maó	kəntávə ~ kəntávə kəntávəs ~ kəntávəs kəntávə ~ kəntávə...	kəntávə kəntávəs kəntávə...

• Alternança entre formes dialectals i l'estàndard

Localitat	<i>LaFV</i>	COD
Tremp	beíβe ~ bée ~ beβíe beíβes ~ bées ~ beβíes beíβe ~ bée ~ beβíe beíβem ~ béem ~ beβíew beíβew ~ béew ~ beβíew beíβen ~ béen ~ beβíen	beβía ~ beía beβíes ~ beíes beβía ~ beíe beβíem ~ beíem beβíew ~ beíew beβíen ~ beíen
el Pont de Suert	bejéβa, bejéβes bejéβe bejéβom bejéβow bejéβen	beβía ~ bejéβa beβíes ~ bejéβes beβíe ~ bejéβe beβíam ~ bejéβom beβíaw ~ bejéβow beβíen ~ bejéβen
Tamarit	perdíβa ~ perdíβo perdíβes perdíβe perdíβem perdíβew perdíβen	perðía ~ perðíβa perðíβes ~ perðíes perðíβe ~ perðíe perðíem ~ perðíβem perðíβew ~ perðíew perðíen ~ perðíβen
Cocentaina	beí <u>a</u> beíes ~ beβíes beía ~ beβía...	beβía ~ beγía ~ beía beβíes ~ beγíes ~ beíes beβía ~ beγía ~ beía...

No hi ha, en aquest temps, cap mostra de tendència a la dialectalització.

4.2.6. Imperfet de subjuntiu

a) Tendència a l'estandardització

- Generalització de la desinència *-i*, llevat del balear, de la vocal temàtica *é*.

Localitat	<i>LaFV</i>	COD
la Seu d'Urgell	kantás ~ kantés ~ kantésə kantásēs ~ kantésēs kantás ~ kantés kantásom kantésow kantásen ~ kantésen	kantés kantésis kantés kantésim kantési kantésin
Andorra	kantésə kantésēs kantés kantésom kantésəw kantésen	kantés kantésis kantés kantésim kantési kantésin
Lleida	perdése ~ perýése perdesēs ~ perýeses perdés ~ perýés perdésēm ~ perýésēm perdésəw ~ perýésəw perdésen ~ perýésen	perðés perðésis perðés / perðési perðésim perðési perðésin
Felanitx	səntís ~ səntijés səntísəs ~ səntísis ~ səntijésəs ~ səntijésis...	səntís səntísis...
Sóller	kəntásəs ~ kəntásis...	kəntásis...
Sóller, Eivissa, Felanitx	pərdésəs ~ pərdésis...	pərdésis...
Eivissa	səntís ~ səntiyés	səntís

- Canvis en la vocal temàtica

Localitat	LaFV	COD
Felanitx	cəntə́s cəntə́səs cəntə́s	cəntás ~ cəntés cəntásis ~ cəntésis cəntás ~ cəntés
Ciutadella	kəntás	kəntés ~ kəntás

- Simplificació del radical

Localitat	LaFV	COD
el Pont de Suert	bejyəsə bejyésés bejyés bejyésom bejyésow bejyésen	beýés ~ beýésa beýésés ~ beýésis beýés beýésim ~ beýésom beýésiwi ~ beýésow beýésen ~ beýésin

- Ús del morf de mode/temps *-si*, que pot conviure o substituir a *-se*.

Localitat	LaFV	COD
Tremp	perýés perýésés perýés perýése perýésew perýésen	perýés ~ perýési perýésés ~ perýésis perýés ~ perýési perýése ~ perýésim perýésew ~ perýésiwi perýésen ~ perýésin

- Convivència de formes, una de les quals pot ser l'estàndard

Benavarri	sentís, sentíse sentíses sentís sentísem sentísew sentísen	sentís ~ sintése ~ sintíse sentíses ~ sintésés ~ sintíses sentís ~ sintése ~ sintíse sentísem ~ sintése ~ sintísem sentísew ~ sintésew ~ sintísiw sentísen ~ sintésen ~ sintísen
-----------	---	---

Gandesa	perýése perýeses perýés perýésem perýésew perýésen	perðésə ~ perðiyésə perðeses ~ perðiyésés perðés ~ perðiyés perðésem ~ perðiyésem perðésew ~ perðiyésew perðésen ~ perðiyésen
Vinaròs	partiýés partiýeses partiýés partiýésem partiýésew partiýésen	serβíra ~ serβiyés ~ serβiskés serβíres ~ serβiyésés ~ serβiskésés serβíra ~ serβiyés ~ serβiskés serβírem ~ serβiyésem ~ serβiskésesem serBírew ~ serβiyésew ~ serβiskésew serBíren ~ serβiyésen ~ serβiskésen

- Reducció de formes alternatives

Localitat	<i>LaFV</i>	COD
Castelló	partíre ~ partiýére partíres ~ partiýéres partíre ~ partiýére...	serβíre serβíres serβíre...
Castelló	seɲkére ~ sentíre seɲkéres ~ sentíres seɲkére ~ sentíre...	sentíre sentíres sentíre...
Eivissa	səntís ~ səntiyés...	səntís...
Manacor	pərtís ~ pərtijés pərtísəs ~ pərtísis ~ pərtijésəs ~ pərtijésis...	sərvís sərvísəs...

b) Tendència a la divergència

- Canvis en el radical

Localitat	<i>LaFV</i>	COD
Falset	beyés beyésis beyés beyésim beyésiwi beyésin	beyés ~ bejyés beyésis ~ bejyésis beyés ~ bejyés beyésim ~ bejyésim beyésiwi ~ bejyésiwi beyésin ~ bejyésin

Lleida	sintiyése sintiyeses sintiyés sintiyésəm sintiyésəw sintiyésən	sentís ~ sentiyés sentísis ~ sentís sentiýés ~ sentísem sentísim ~ sentiýésim sentísew ~ sentísiw ~ sentiýési sentísen ~ sentísin ~ sentiýésin
Vinaròs	beýés beýeses beýés	beýés ~ bejýés beýeses ~ bejýeses beýés ~ bejýés

- Manteniment de les formes velaritzades

Localitat	LaFV	COD
Sagunt	perdéra ~ perdéra ~ perýéra ~ perýérə perdéres ~ perýéres perdéra ~ perýéra ~ perýérə	perýéra perýéres perýéra
Formentera	pərtísəs ~ pərtísis ~ pərtiyésəs ~ pərtiyésis	servísis ~ sərviyésis
Eivissa	pərtiyésəm ~ pərtiyésim	sərviyésim

4.2.7. El cas del pretèrit perfet

L'evolució del perfet es pot examinar des de dues perspectives: a) en relació amb la velarització del paradigma que afecta els verbs del tipus *perdre*, un fet prou freqüent; b) en relació amb la substitució del perfet simple pel perifràstic.

a)

Localitat	LaFV	COD
Sagunt Sueca València	perýí perýéres perýé perýérem perýérew perýéren	perýí perýéres perýé perýérem perýérew perýéren
Llíria	perýí perýéres perýé...	perðí ~ perýí perðéres ~ perýéres perðé ~ perýé...

b) En valencià i en balear s'ha tendit a substituir el perfet simple, que documentava *LaFV*, pel perfet perifràstic, que il·lustra el COD.

En general el perfet es manté a Llúria, Sueca i València (però no en tots els verbs examinats). Es canvia la forma simple per la perifràstica a Sagunt en certs verbs, cosa que indica que la substitució no acaba de ser general. En alguns casos les formes perifràstiques s'alternen amb les formes simples d'algunes persones. A Sueca no es documenta, en el COD, l'ús de la primera persona del singular. A Elx i a Sagunt no utilitzen ni la primera ni la tercera persona del singular.

En la varietat balear, el perfet simple és reemplaçat pel perifràstic a Pollença, Sóller, Manacor i Palma (que poden conservar escadusserament en alguns verbs les formes simples en les persones del plural); el mateix succeeix a Maó, Felanitx, Eivissa i Formentera, que ja no utilitzaven, segons *LaFV* la primera persona del singular del perfet simple.

4.2.8. Altres canvis: el cas de les vocals temàtiques

S'observa que en les dades del COD es conserven molt escadusserament les vocals temàtiques etimològiques en el present i l'imperatiu (persones 4 i 5) en alguns indrets:

- En la primera persona del plural de l'imperatiu

Benavarri	<i>LaFV</i> [perdám]	COD [perðám ~ perðém]
Benavarri	<i>LaFV</i> [beywém]	COD [beyám ~ beywém]
Benavarri	<i>LaFV</i> [partím]	COD [sirβám]
Tortosa	<i>LaFV</i> [kantám ~ kantém]	COD [kantém]

- En la segona persona del plural de l'imperatiu

Benavarri, el Pont de Suert i Tortosa	<i>LaFV</i> [kantáw ~ kantéw]	COD [kantéw]
Benavarri	<i>LaFV</i> [perdáw]	COD [perðáw ~ perðéw]
Benavarri	<i>LaFV</i> [perdáw]	COD [perðáw ~ perðéw]
Benavarri	<i>LaFV</i> [partíw]	COD [sirβáw]

4.2.9. Altres canvis: els subjuntius acabats en -o

LaFV registra el present de subjuntiu acabat en -o, de vegades en forma alternada, a Gandesa ([béyes ~ béyos] i al COD: [béyes ~ béjyos]). A Lleida, Balaguer, Andorra, la Seu d'Urgell i Morella també apareix la mateixa alternança, però ja no es recull en el COD. També es troba en la segona persona de l'imperatiu a Tamarit: [kánto] (però al COD: [kánte ~ kánti]). La GBU (p. 81) considera que les formes en -os són pròpies dels registres informals i s'han d'evitar en les situacions de formalitat.

5. QUANTIFICACIÓ DE LES DIFERÈNCIES: CAP A UN ANIVELLAMENT DIALECTAL EN LA FLEXIÓ VERBAL REGULAR?

Per valorar i quantificar les diferències de les formes verbals regulars de cada localitat s'ha pres com a referència el que determinen els tractats normatius: d'una banda, la *Proposta per a un estàndard oral de la llengua catalana II. Morfologia*; de l'altra, la gramàtica de l'Institut d'Estudis Catalans en les seves dues versions (GIEC i GBU). Hi ha relatives petites diferències entre totes dues fonts. La primera incorporava les etiquetes dels àmbits (restringit – general), no gaire afortunades, i de l'acceptació normativa (admissible – no recomanable), les quals es podien combinar des d'un punt de vista diafàsic (registre informal – molt formal). La darrera obra, la voluntat de la qual és ser tant descriptiva com normativa, incorpora els conceptes anteriors eliminant-ne els àmbits. Quant als usos lingüístics, manté la dicotomia entre «(molt) formal» i «informal»; i, quant a l'ús normatiu, utilitza les etiquetes «no acceptable», «s'evita» i «és preferible».

Prenent com a referència les formes acceptades en cada àrea dialectal d'acord amb la normativa, el recompte de les discrepàncies s'ha dut a terme mitjançant una valoració binària (0 = coincidència; 1 = divergència) que s'ha aplicat a les formes verbals en relació amb l'estàndard.

El recompte de les divergències s'ha dut a terme des de la perspectiva de tres variables: a) l'àrea dialectal; b) el temps verbal; c) la localitat.

Un cop determinats els percentatges de divergència per a cada àrea dialectal, s'ha fet la comparació entre els dos repertoris, de manera que, en relació amb les variables esmentades, es pot observar quina mena de variació s'ha produït. La variació pot ser positiva, en el sentit que s'ha incrementat el nombre de divergències en relació amb l'estàndard, atès que hi ha hagut un aug-

ment de punts percentuals; o, a la inversa, la variació pot ser negativa, en el sentit que ha disminuït el nombre de divergències en relació amb l'estàndard, per raó que hi ha hagut una minva de punts percentuals.

5.1. Les àrees dialectals

Es presenten a continuació els resultats globals relatius a les tres àrees dialectals examinades:

a) Occidental

LaFV: d'un total de 3.315 formes; 0 = 2.243; 1 = 972. Les divergències en relació amb l'estàndard representen el 41,48 %.

COD: d'un total de 4.947 formes; 0 = 4026; 1 = 921. Les divergències en relació amb l'estàndard representen el 18,61 %.

Si es comparen els dos resultats, hi ha, doncs, una reducció de les divergències en 22,87 punts percentuals.

b) València

LaFV: d'un total de 3.309 formes; 0 = 2.991; 1 = 518. Les divergències en relació amb l'estàndard representen el 15,65 %.

COD: d'un total de 4.912 formes; 0 = 4.147; 1 = 765. Les divergències en relació amb l'estàndard representen el 15,57 %.

Si es comparen els dos resultats, hi ha, doncs, una reducció de les divergències en 0,08 punts percentuals.

c) Balear

LaFV: d'un total de 2.580 formes; 0 = 2.243; 1 = 337. Les divergències en relació amb l'estàndard representen el 13,06 %.

COD: d'un total de 2.946 formes; 0 = 2.772; 1 = 174. Les divergències en relació amb l'estàndard representen el 5,9 %.

Si es comparen els dos resultats, hi ha, doncs, una reducció de les divergències en 7,16 punts percentuals.

La taula següent sintetitza els percentatges de divergències de manera global en els tres dialectes i, en tots els casos, la variació negativa que s'ha produït si es comparen els resultats de tots dos repertoris:

	Occidental	Valencià	Balear
LaFV	41,48 %	15,65 %	13,06 %
COD	18,61 %	15,57 %	5,9 %
Variació negativa	↓ 22,87	↓ 0,08	↓ 7,16

La taula indica que en l'àrea occidental, d'acord amb les dades de *LaFV*, és on es produeix el nombre més elevat de divergències en relació amb les realitzacions de l'estàndard. Les varietats valenciana i balear també mostren divergències pel que fa als paradigmes regulars normatius, però en un percentatge molt inferior.

Les dades del COD ofereixen el mateix panorama quant a la gradació, però els resultats de les divergències relatives a l'estàndard són molt inferiors. Els percentatges, doncs, registren que en el període temporal de tres generacions s'han reduït considerablement les divergències en relació amb l'estàndard i que, en contrapartida, s'ha produït un anivellament en les formes verbals regulars. En relació amb els resultats percentuals, si es contrasten les dades dels dos repertoris s'aprecia que on més s'ha reduït el nombre de divergències és en l'àrea occidental. La reducció en l'àrea balear no arriba a 10 punts percentuals i és insignificant en la varietat valenciana.

5.2. Els temps i modes verbals

Des del punt de vista dels temps verbals, les dades dels dos projectes mostren més divergències en relació amb les solucions de l'estàndard i en quin percentatge. S'ha deixat de banda la comptabilització de la incidència de les formes no estàndard en les classes verbals per no allargar excessivament aquest treball, però es pot hipotetitzar que la primera conjugació, que és la més estable, seria la que presentaria un nombre menys elevat de divergències.

	Occidental <i>LaFV</i>	Occidental COD	València <i>LaFV</i>	València COD	Balear <i>LaFV</i>	Balear COD
Condicional	20,13 %	11,02 %	1,63 %	1,75 %	0	0
Futur	0	1,23 %	1,49 %	2,46 %	0	0
Gerundi	16,66 %	20,25 %	1,35 %	16,16 %	12,23 %	23,8 %
Imperfet d'indicatiu	48,46 %	22,77 %	6,18 %	9,92 %	6,18 %	0

Imperatiu	23,24 %	28,74 %	19,29 %	32,83 %	7,09 %	18,62 %
Infinitiu	8,45 %	0	1,44 %	2,27 %	0	0
Imperfet de subjuntiu	48,61 %	46,64 %	26,33 %	24,82 %	28,84 %	12,44 %
Participi	1,41 %	0	7,14 %	3,79 %	5,88 %	0
Perfet	0	0	19,20 %	7,76 %	4,91 %	0
Present d'indicatiu	20,04 %	9,9 %	11,11 %	11,16 %	1,42 %	0
Present de subjuntiu	43,30 %	31,97 %	41,71 %	34,89 %	28,79 %	13,42 %

Les dades de *LaFV* reflecteixen que en l'àrea occidental les formes de futur i de perfet no presenten cap divergència en relació amb l'estàndard. Les divergències de les formes impersonals no arriben al 10%, llevat del gerundi, que supera el 15%. Les formes de present d'indicatiu, de condicional i d'imperatiu es troben en un interval entre el 20% i el 25%. Finament, el present de subjuntiu, l'imperfet d'indicatiu i l'imperfet de subjuntiu sobrepassen el 40% de divergències. A grans trets, aquestes divergències es poden atribuir al polimorfisme que presenten aquests temps i a l'expansió de la velarització.

En relació amb les divergències que presenten els temps verbals en l'àrea valenciana, d'acord amb les dades de *LaFV*, el panorama experimenta un cert canvi. No hi ha cap temps que no presenti divergències, més o menys considerables, en relació amb l'estàndard. En ordre creixent, les formes impersonals, el futur, el condicional i l'imperfet d'indicatiu no superen ni de lluny el 10% de divergències. El present d'indicatiu sobrepassa aquesta xifra per molt poca diferència i el perfet no arriba gairebé al 20%, la qual cosa es pot explicar igualment per l'expansió de la velarització i també pel polimorfisme. Els temps de subjuntiu són els que presenten solucions més discrepants: l'imperfet de subjuntiu en un grau molt menor que el present de subjuntiu, atès que la forma gairebé general en una bona part del domini, *-ra*, és susceptible, però, d'aparèixer en totes les conjugacions II i III, que no els pertocaria d'experimentar una velarització. El present de subjuntiu és el temps que presenta més divergències, en una xifra que s'acosta a la que mostrava el català occidental.

Finalment, els temps verbals del balear recollits en *LaFV* mostren que sols comparteixen alguns trets comuns si es comparen amb les varietats dialectals anteriors. D'una banda, ni l'infinitiu ni els temps que històricament se'n deriven (el futur i el condicional) no presenten cap divergència en relació amb l'estàndard. La resta, llevat dels temps de subjuntiu, presenten una incidència de formes divergents relativament baixa. Cap no supera el 10%, excepte les

formes de gerundi, susceptibles de velaritzar-se. El present i l'imperfet de subjuntiu mostren un percentatge similar de divergències, sols una mica superior en aquest darrer temps. En cap cas, però, els resultats s'acosten a les xifres presentades per l'occidental i el valencià.

Quant als resultats del COD, en l'àrea occidental l'ordenació dels temps pel que fa al nombre de divergències en relació amb l'estàndard presenta certes similituds amb el que s'ha observat en les dades de *LaFV*. En aquest cas, l'infinitiu, el participi i el perfet tampoc no mostren divergències. El futur i el present d'indicatiu mostren un percentatge de divergències que no arriba al 10% i el condicional se'n separa en un 11%. El gerundi, l'imperfet d'indicatiu i l'imperatiu presenten un interval de divergències que oscil·la entre el 20% i el 30%. Els temps de subjuntiu, tot i presentar divergències, tenen percentatges prou allunyats, i és l'imperfet de subjuntiu, un cop més, i com en el repertori anterior, el temps amb formes més allunyades de l'estàndard (un 46%).

Novament, el panorama que ofereixen les varietats valencianes, d'acord amb les dades del COD, canvia prou si es comparen amb les de l'àrea occidental. En aquest cas, tots els temps presenten un percentatge més o menys elevat de divergències en relació amb l'estàndard. Uns quants (l'infinitiu, el futur, el condicional, el participi, el perfet i l'imperfet d'indicatiu), en un percentatge que no supera el 10%. Les divergències en les formes de present d'indicatiu i de gerundi no arriben en cap cas al 20%. L'imperfet de subjuntiu, a diferència del mateix temps en occidental, no supera el 25% de divergències. L'imperatiu i el present de subjuntiu són els temps amb més formes divergents, les quals no superen mai, però, el 35%.

Finalment, les dades del COD recollides en les localitats balears presenten una panoràmica ben diferent del que s'observava en les àrees dialectals anteriors. Un bon nombre de temps (el condicional, el futur, l'imperfet d'indicatiu, el perfet i el present d'indicatiu) i de modes (l'infinitiu i el participi) no mostren cap divergència en relació amb les solucions de l'estàndard, amb la qual cosa és evident que l'anivellament dialectal ja s'ha produït. Sorprenentment, contrastant-les amb les dades anteriors, ni els temps de subjuntiu ni el mode imperatiu presenten el percentatge més elevat de discrepàncies (en tots els casos no supera el 20%). El mode que mostra més divergències, també atribuïbles en aquest cas a la velarització, és el gerundi, en un percentatge superior al que mostraven les dades equivalents de *LaFV*.

Contraposats els resultats de tots dos repertoris en relació amb les divergències que presenten els temps i els modes verbals respecte a l'estàndard, s'observa una variació positiva, en termes de punts percentuals, segons si s'han incrementat les divergències en relació amb l'estàndard, o bé una variació negativa, segons si aquestes divergències han disminuït.

Els dos tipus de variació esmentada es poden veure en la taula següent (amb un fons gris i amb la fletxa cap amunt, els temps que han experimentat una variació positiva; és a dir, un increment de diferències en relació amb les solucions normatives).

	Occidental	Valencià	Balear
Condicional	↓ 9,11	↑ 0,12	—
Futur	↑ 1,23	↑ 0,99	—
Gerundi	↑ 3,59	↑ 16,16	↑ 11,57
Imperfet d'indicatiu	↓ 25,69	↑ 3,74	↓ 6,18
Imperatiu	↑ 5,5	↑ 13,54	↑ 18,62
Infinitiu	↓ 8,45	↑ 0,83	—
Imperfet de subjuntiu	↓ 1,97	↓ 1,51	↓ 16,4
Participi	↓ 1,41	↓ 3,35	↓ 5,88
Perfet	—	↓ 11,44	↓ 4,91
Present d'indicatiu	↓ 10,14	↑ 0,05	↓ 1,42
Present de subjuntiu	↓ 11,33	↓ 6,82	↓ 15,37

De manera global, el primer que s'observa és que, en relació amb l'increment en el nombre de discrepàncies, els temps i els modes varien d'acord amb l'àrea dialectal, i és en valencià on n'apareix un nombre més elevat. Amb tot, cal dir que, malgrat el nombre de temps i modes afectats, les xifres d'augment en punts percentuals són en realitat prou baixes: des de la irrellevant pujada de l'infinitiu, del futur, del condicional i del present d'indicatiu fins a les altres formes de gerundi i d'imperatiu, que no superen el 20%.

En occidental, s'han incrementat les divergències en el gerundi i l'imperatiu, cosa que d'alguna manera podia ser esperable. El lleuger increment del futur respon al fet que en *LaFV* aquest temps no presentava cap mena de divergència, ni en el radical ni en les desinències.

En darrer terme, en la varietat balear són precisament també el gerundi i l'imperatiu els més afectats, relativament (observem que no superen el 12%), en relació amb l'augment en les divergències pel que fa a l'estàndard.

Malgrat les diferències en el nombre i el lloc de les discrepàncies, crec que és possible donar una explicació conjunta que les justifiqui.

5.3. Un apunt sobre les possibles causes de les divergències dialectals

Ja s'ha vist que són les formes de subjuntiu i, en alguna varietat, també les d'imperatiu les que mostren més divergències en relació amb les solucions normatives. I, pel que fa a les causes, ja s'ha apuntat al polimorfisme i a la tendència a la velarització. És cert que l'anivellament dialectal implica una reducció de formes alternatives a favor de l'estàndard, però malgrat la seva «prohibició», la velarització ha avançat i avança de manera inexorable. I un dels motors que poden justificar aquesta tendència és l'analogia.

Així, el present de subjuntiu és un dels temps, juntament amb l'imperfet de subjuntiu, que presenta més divergències en relació amb l'estàndard. I una de les causes és l'expansió del segment velar tant intraparadigmàticament (a verbs del tipus *perdre* i a formes del verb *beure* en les quals aquest segment no és admès: *beguent*, *dolguent*, *molguent* i, encara, *perguent*) com interparadigmàticament, és a dir, a verbs de la primera i la tercera conjugacions.

A més, el segment velar /g/, que presenta inicialment l'al·lofonia [ɣ] ~ [k], pot adquirir també altres aparences o variants: [iɣ] o [eɣ]. Wheeler (1985: 564) considera que [iɣ] és una variant sil·làbica de /g/, marca de subjuntiu, que evita les dificultats articulatòries quan s'adjunta a una arrel acabada en determinades consonants (que no siguin vocal o /ʎ/, /m/ o /r/).

Així, l'increment de caràcter velar es difon, amb variants formals diferents —/g/, /ig/, /eg/—, per acomplir la seva funció de marcador de determinades categories morfosintàctiques: tant subclasse verbal com mode/temps. L'adopció d'una d'aquestes variants depèn que l'arrel sigui susceptible de modificar-se o no, en funció de la varietat dialectal, davant la presència de l'increment velar. Si es pot alterar, adopta la solució velar [g] ([bágga], [dórrɣga]); si no ho pot fer, s'aplica, en funció dels dialectes, la forma [iɣ] ([bátiɣa], [dórmiga]) (o [eɣ]: [báteɣa], [dórmeyɣa]), i els resultats esdevenen esdrúixols en algunes persones. A més dels exemples adduïts en relació amb la segona i la tercera conjugacions, ja s'ha vist que el segment velar s'expandeix també a temps de la primera conjugació:

Present de subjuntiu

<i>LaFV</i>	COD	<i>LaFV</i>	COD
el Pont de Suert	el Pont de Suert	Balears	Balears
cánt + ey + a	cánt + ey + a	cánt + iy + a	cánt + i
cánt + ey + e + s	cánt + ey + e + s	cánt + iy + e + s	cánt + i + s
cánt + ey + a	cánt + ey + a	cánt + iy + a	cánt + i
cant + ey + é + m			
cant + ey + é + w			
cánt + ey + e + n	cánt + ey + e + n	cant + iy + e + n	cant + i + n

Aquesta expansió en el present de subjuntiu també pot afectar l'imperfet de subjuntiu.

el Pont de Suert	Tortosa
cant + é + sa / cant + ey + é + sa	cant + é + s / cant + iy + é + sa / -é + s
cant + é + se + s / cant + ey + é + se + s	cant + é + se + s / cant + iy + é + se + s
cant + é + s / cant + ey + é + s	cant + é + s / cant + iy + é + sa / -é + s
cant + é + so + m / cant + ey + é + so + m	cant + é + se + m / cant + iy + é + se + m
cant + é + so + w / cant + ey + é + so + w	cant + é + se + w / cant + iy + é + se + w
cant + é + se + n / cant + ey + é + se + n	cant + é + se + n / cant + iy + é + se + n

A més, s'ha omès voluntàriament, perquè s'allunya de la finalitat d'aquest estudi, la presència en el paradigma verbal de marques dobles com ara *mò + g + uig + a* (Tortosa) (estàndard, *mogui*), *part + èixqu + igu + es* (Eivissa) (estàndard, *parteixis*).

La difusió del segment velar en les varietats dialectals examinades es pot explicar, a grans trets, per tres processos que l'afavoreixen:

- a) L'analogia:⁸ l'exemple següent en pot ser una mostra. Així, d'acord amb la GBU, p. 83, «[en] parlars baleàrics (i, en menor mesura, també en alguns parlars nord-occidentals)» [...] «són acceptables, tot i que tenen menys extensió d'ús, les formes amb el segment amb velar *-igu-* de la primera i segona persona del plural del present de subjuntiu dels verbs de la tercera conjugació: 1pl *dormiguem*, 2pl *dormigueu*». A partir d'aquestes formes admeses és molt fàcil, si no provenen d'altres temps, la propagació a altres persones per raó de l'analogia.

8 Vegeu una caracterització del concepte d'analogia a Elvira (1998).

- b) L'elusió del sincretisme o identitat formal: en el conjunt de les realitzacions verbals regulars del domini català hi ha temps que admeten més sincretisme que d'altres, o varietats dialectals més susceptibles de manifestar-lo en alguns punts del paradigma. Com a exemple, observem que la presència del segment velar en les dades del present d'indicatiu del verb *cantar* al Pont de Suert podria justificar-se per evitar formes sincrètiques, que són admeses, però, en la majoria de dialectes. A partir de la presència en la primera i la segona persones del plural, el segment velar podria estendre's, per analogia, a la resta del paradigma:

1	cant + o	cánt + ey + a	↑
2	cant + e + s	cánt + ey + e + s	↑
3	cant + a	cánt + ey + a	↑
4	cant + a + m / cant + e + m	⇒ cant + ey + e + m	
5	cant + a + w / cant + e + w	⇒ cant + ey + e + w	
6	cant + e + n	cánt + ey + e + n	↓

- c) Motivacions psicolingüístiques. Els parlants sempre cerquen transparència i tendeixen a rebutjar l'opacitat. Adoptant el parer de Wheeler (1985: 561), desenvolupat en el marc de la morfologia natural, es poden justificar els canvis morfològics motivats per raons psicolingüístiques i comunicatives. En efecte, certs punts «febles», des del punt de vista de la transparència, en el sistema morfològic verbal poden produir problemes psicològics o psicolingüístics en el parlant a causa de l'ambigüïtat que poden crear. La tendència a evitar o a reduir la sinonímia o l'ambigüïtat pot provocar, certament, innovacions lingüístiques.

5.4. Les localitats

També en el cas de les diverses localitats s'han pogut comptabilitzar, en comparar les dades de tots dos repertoris, les que han experimentat una variació negativa; és a dir, aquelles en les quals, per punts percentuals, han disminuït les divergències en relació amb l'estàndard. Aquestes localitats són, en les tres àrees dialectals examinades, la majoria.

Al seu torn, també s'han pogut determinar les que experimenten una variació positiva; és a dir, on s'han incrementat les divergències en relació amb les formes estàndard.

En la taula següent, en l'àrea occidental augmenten les divergències a Benavarri, Fraga i Falset, tot i que amb uns valors relativament baixos, que no superen en cap cas els 14 punts.

En contrapartida, un nombre de localitats redueixen les divergències en relació amb la varietat estàndard de manera prou notable. Cal observar els resultats d'Andorra, Lleida i Tremp, les divergències de les quals s'incrementen en més de 20 punts; i a la Seu d'Urgell en 34 punts.

Localitat	LaFV	COD	Variació
Andorra	33,8 %	9,89 %	↓ 23,91
Balaguer	29,72 %	22,10 %	↓ 7,62
Benavarri	28,18 %	41,31 %	↑ 13,13
Cervera	11,66 %	4,33 %	↓ 7,33
el Pont de Suert	41,55 %	29,29 %	↓ 12,26
Falset	0,45 %	9,81 %	↑ 9,36
Fraga	13,21 %	19,06 %	↑ 5,85
Gandesa	26,97 %	25 %	↓ 1,97
la Seu d'Urgell	37,99 %	3,56 %	↓ 34,43
Lleida	35,09 %	10,61%	↓ 24,48
Sort	43,20 %	32,36%	↓ 10,84
Tamarit	38,22 %	32,13%	↓ 6,09
Tortosa	30,08 %	25,51%	↓ 4,57
Tremp	47,38 %	21,5%	↓ 25,88

Pel que fa a les localitats valencianes, Alcoi, Sueca i Vinaròs mostren un increment en les divergències, sense superar, però, els 10 punts. I a Alacant aquesta xifra augmenta una mica més. La reducció de les divergències en les altres localitats no és especialment significativa. Sols a Castelló no arriba per poc als 12 punts.

Alacant	9,09 %	20,32%	↑ 11,23
Alcoi	6,66 %	16,44%	↑ 9,78
Castelló	28,18 %	16,22%	↓ 11,96
Cocentaina	19,65 %	12,66%	↓ 6,99
Elx	9,54 %	9,86%	↑ 0,32
Gandia	13,88 %	11,87%	↓ 2,01
la Vila Joiosa	15,28 %	13,93%	↓ 1,35
Llíria	18,08 %	13,74%	↓ 4,34
Morella	23,88 %	20,72%	↓ 3,16
Sagunt	15,03 %	13,66%	↓ 1,37
Sueca	16,72 %	21,13%	↑ 4,41
València	18,59 %	14,05%	↓ 4,54
Vinaròs	16,03 %	18,76%	↑ 2,73

Finalment, en l'àrea balear predomina, per bé que amb una xifra no gaire significativa, llevat de Manacor i Felanitx, que superen els 10 punts, la reducció de les diferències en relació amb l'estàndard. Sols Maó experimenta una variació positiva, amb un increment també prou baix: de 4 punts.

Ciutadella	9,17 %	9,12%	↓ 0,05
Eivissa	24,43 %	4,7%	↓ 19,73
Felanitx	14,28 %	4,24%	↓ 10,04
Formentera	12,45 %	6,62%	↓ 5,83
Manacor	14,05 %	2,07%	↓ 11,98
Maó	7,89 %	12,11%	↑ 4,22
Palma	10,3 %	3,6%	↓ 6,7
Pollença	10,13 %	4,88%	↓ 5,25
Sóller	10,34 %	7,44%	↓ 2,9

Aquestes variacions positives s'expliquen bàsicament per la pèrdua de formes alternatives i la incorporació de certs morfs que concorden amb els de l'estàndard. Recordem, per exemple, la difusió dels subjuntius en *-i* i del morf de mode/temps de l'imperfet de subjuntiu *-és* a les Illes.

6. CONCLUSIONS

Ja s'ha vist que la morfologia verbal ofereix un terreny extraordinari per donar compte de fenòmens de canvi lingüístic, els quals desemboquen en un anivellament dialectal. La comparació de les dades ha permès de detectar la naturalesa dels canvis produïts en l'interval de noranta anys. Aquests poden prendre, com s'ha vist, dues direccions: o bé vers l'estandardització, la qual s'ha pogut comptabilitzar des de diverses perspectives (àrees, temps i modes i localitats), o bé vers la dialectalització. Sens dubte, ja s'ha observat que la primera direcció és la predominant.

En aquest cas, la inexorabilitat del canvi és afavorida per altres factors que poden accelerar l'estandardització i que són ben coneguts: a més de l'ensenyament i els mitjans de comunicació, cal valorar el prestigi que té l'estàndard. Paralelament, també s'ha observat que fenòmens basats en l'analogia i en l'elusió del sincretisme provoquen un allunyament de les formes normatives, i el fenomen de l'expansió de la velarització és un dels factors més significatius. És cert que el parlant cerca transparències i eludeix dificultats: només cal veure els recursos basats en les assimilacions, en les dissimilacions o en l'adjunció d'elements epentètics, que desfiguren fins a un cert punt les formes normatives, però que resulten econòmics i prou rendibles.

Els processos de canvi que s'han mostrat no poden respondre la pregunta que Aitchinson planteja: si el canvi empeny la llengua cap a un empobriment o cap a un enriquiment. De fet, el canvi lingüístic no motivat és absolutament contingent i força lluny de ser predictable. En canvi, el canvi controlat, com pot ser el motivat per un procés d'estandardització, significa generar un canvi necessari, que, en el marc de la morfologia verbal, té la finalitat d'uniformar paradigmes, eliminar formes que, de manera arbitrària, es condemnin. Aquesta acció purificadora dona lloc a una reducció de les diferències i, per tant, no pot ser vista com un enriquiment, ans al contrari.

Si aquesta reducció de la variació morfològica ha tingut lloc en un interval de tres generacions, es pot preveure que en tres generacions més els paradigmes verbals de totes les varietats del català esdevindran encara més homogenis. Aleshores no serà possible analitzar la difusió del segment velar ni l'acció de

fenòmens d'evolució analògica. Potser ara és el moment de fer una recopilació de dades verbals que inclogui totes les localitats i tots els verbs que Alcover va sotmetre a enquesta i que tingui en compte parlants de diferents edats. Des d'aquesta perspectiva, la dialectologia desenvoluparà la seva faceta arqueològica.

REFERÈNCIES BIBLIOGRÀFIQUES

- AITCHISON, Jean (2013). *Language change: Progress or decay?* Cambridge: Cambridge University Press.
- ALCOVER, Antoni Maria; MOLL, Francesc de B. (1929-1933). *La flexió verbal en els dialectes catalans*. Barcelona: Anuari de l'Oficina Romànica (vol. II (1929) [73] 1- [184] 112; vol. III (1930) [73] 1- [168] 96; vol. IV (1931) [9] 1- [104] 96; vol. V (1932) [9] 2 - [72] 64). Consultable a alcover.iec.cat.
- AUER, Peter (2018). «Dialect change in Europe - leveling and convergence». A: BOBERG, Charles; NERBONNE, John; WATT, Dominic. *The handbook of dialectology*. Oxford: Wiley Blackwell, pàg. 159-176.
- BIBILONI, Gabriel (2016). *El català de Mallorca: La fonètica*. Palma: Lleonard Muntaner.
- BRITAIN, David (2010). «Supralocal regional dialect levelling». A: LLAMAS, Carmen; WATT, Dominic (eds.). *Language and identities*. Edimburg: Edinburgh University Press, pàg. 193-204.
- GRIMALT, Josep (1972). «El vocalisme de Felanitx». *Mayurqa*, núm. 8, pàg. 57-64.
- ELVIRA, Javier (1998). *El cambio analógico*. Madrid: Gredos.
- HEERINGA, Wilbert; NERBONNE, John (2001). «Change, convergence and divergence among Dutch and Frisian». A: BOERSMA, Piter; BREUKER, Pieter; JANSMA, Lambert; VAART, Jacob van der (eds.). *Philologia frisica anno 1999: Lêzingen fan it fyftjinde Frysk filologekongres*. Ljouwert: Fryske Akademy, pàg. 88-109.
- INSTITUT D'ESTUDIS CATALANS (1999). *Proposta per a un estàndard oral de la llengua catalana I. Morfologia*. Barcelona: Institut d'Estudis Catalans.
- INSTITUT D'ESTUDIS CATALANS (2016). *Gramàtica de la llengua catalana*. Barcelona: Institut d'Estudis Catalans.
- INSTITUT D'ESTUDIS CATALANS (2019). *Gramàtica bàsica i d'ús de la llengua catalana*. Barcelona: Empúries / Institut d'Estudis Catalans.
- KABATEK, Johannes; PUSCH, Claus D. (eds.). «Variació, estandardització, globalització: fenòmens de convergència i divergència lingüístiques. Introducció a la temàtica del volum». A: *Variació, poliglòssia i estàndard*. Aachen: Shaker, pàg. 1-19.
- KERSWILL, Paul (2003). «Dialect levelling and geographical diffusion in British English». A: BRITAIN, David; CHESHIRE, Jenny. *Social dialectology in honour of Peter Trudgill*. Amsterdam / Filadèlfia: John Benjamins, pàg. 223-234.

- LLADÓ, Jordi; BORRÀS-COMES, Joan (2012). «Morfologia verbal i anivellament dialectal: el cas del present de subjuntiu calero». *BeCeroLes*, núm. 5, pàg. 79-85.
- LLORET, Maria-Rosa; PEREA, Maria Pilar (2002). «A report on "The Corpus Oral Dialectal del Català Actual" (COD)». *Dialectologia et Geolinguística*, núm. 10, pàg. 59-76.
- MALKIEL, Yakov (1957-1958). «Diachronic hypercharacterization in Romance». *Archivum Linguisticum*, núm. IX, pàg. 79-113, i núm. X, pàg. 1-36.
- MASSANELL, Mar (2001). «Morfologia flexiva actual de la Seu d'Urgell i Coll de Nargó: estudis en el procés d'orientalització del català nord-occidental». *Zeitschrift für Katalanistik: Revista d'Estudis Catalans*, núm. 14, pàg. 128-150.
- MASSANELL, Mar (2012). «*Fewe temps que no diva tants verbs!*»: *Manteniment i transformació de paradigmes verbals en el català nord-occidental del tombant de segle*. Barcelona: Publicacions de l'Abadia de Montserrat.
- MOLL, Francesc de B. (1991). *Gramàtica històrica catalana*. València: Universitat de València.
- MOLL, Francesc de B. (2006). *Gramàtica històrica catalana*. València: Universitat de València.
- PEREA, Maria Pilar (1999). *La flexió verbal en els dialectes catalans d'A. M. Alcover i F. de B. Moll: Compleció i ordenació*. 2 vols. Barcelona: IEC.
- PEREA, Maria Pilar (2001). *La flexió verbal en els dialectes catalans d'A. M. Alcover i F. de B. Moll: Les dades i els mapes*. Palma: Conselleria d'Educació i Cultura, Govern de les Illes Balears, edició en CD-ROM.
- PEREA, Maria Pilar (2005a). *Dades dialectals. Antoni M. Alcover*. Palma: Conselleria d'Educació i Cultura, Govern de les Illes Balears, edició en CD-ROM.
- PEREA, Maria Pilar (2005b). «La flexió verbal en els dialectes catalans». A: *Antoni M. Alcover dialectòleg, gramàtic, polemista*. Barcelona: Publicacions de l'Abadia de Montserrat, pàg. 117-147.
- PEREA, Maria Pilar (2007a). «Dedialectalization or the death of a dialect: the case of the catalan subdialect spoken in the Costa Brava». *Dialectologia et Geolinguística*, núm. 15, pàg. 77-89.
- PEREA, Maria Pilar (2007b). «Phonetic and morphological variation in "Rossellonès"». *Studies in Eurolinguistics* (Berlín: Logos), vol. 5, pàg. 199-209.
- PEREA, Maria Pilar (2009a). «Elements exolingües en el català del Rosselló». A: KABATEK, Johannes; PUSCH, Claus D. (eds.). *Variació, poliglòssia i estàndard*. Aachen: Shaker, pàg. 85-102.
- PEREA, Maria Pilar (2010). «The dialect of Alghero: continuity and change». A: MILLAR, Robert M. (ed.). *Marginal dialects: Scotland, Ireland and Beyond*. Aberdeen: University of Aberdeen, pàg. 131-148.
- RADATZ, Hans Ingo (2007). «La desdialectalització del català de Mallorca». *Els Marges*, núm. 81, pàg. 29-41.
- VALLS, Esteve (2008). *La desdialectalització del català nord-occidental: cap a una convergència total amb l'estàndard? Una anàlisi dialectomètrica sobre el procés d'anivella-*

- ment dialectal en quatre generacions de parlants nord-occidentals*. Treball d'investigació inèdit.
- VALLS, Esteve (2013). «Canvi morfològic vs. canvi fonològic en català nord-occidental». *Treballs de Sociolingüística Catalana*, núm. 23, pàg. 57-79.
- VALLS, Esteve (2017). «Cap a on va el català de la Franja? Alguns exemples de canvi lingüístic en curs». A: GIRALT, Javier; MORET, Maria Teresa. *El repte d'investigar sobre la Franja d'Aragó: Jornada de l'Associació Internacional de Llengua i Literatura Catalanes a Saragossa (28 d'octubre de 2016)*. Saragossa: Prensas de la Universidad de Zaragoza, pàg. 51-86.
- VALLS, Esteve (2019). *La llengua escapçada: Un estudi sobre l'autonomització dels parlars nord-occidentals a la cruïlla catalanoaragonesa*. Barcelona: Generalitat de Catalunya, Departament de Cultura, Direcció General de Política Lingüística.
- VIAPLANA, Joaquim (1986). «Morfològia flexiva i flexió verbal catalana». *Llengua & Literatura* (Barcelona: IEC), núm. 1, pàg. 385-403.
- VIAPLANA, Joaquim; LLORET, Maria-Rosa; PEREA, Maria-Pilar; CLUA, Esteve (2007). *Corpus oral dialectal*. Barcelona: Universitat de Barcelona, publicació en CD-ROM.
- WHEELER, M. W. (1985) «Analogia i psicologia: el desenvolupament de la morfològia verbal balear». A: *Actes del XVI Congrés Internacional de Lingüística i Filologia Romàniques*. Vol. II. Palma: Moll, pàg. 557-566. [Publicat amb ampliacions sota el títol «Analogy and psychology: morphological change in Balearic Catalan» a *Sheffield Working Papers in Language & Linguistics*, núm. 3 (1986), pàg. 1-16.]

Secció de Lingüística Catalana
Departament de Filologia Catalana



 UNIVERSITAT DE
BARCELONA
Edicions

COL·LECCIÓ
LINGÜÍSTICA
CATALANA

22