Viaplana, Joaquim; Maria-Rosa Lloret, Maria-Pilar Perea i Esteve Clua (2007): COD. Corpus Oral Dialectal. Barcelona: Promociones y Publicaciones Universitarias (PPU). [CD-ROM]
ISBN: 978-84-477-0990-8

Criteris

Aquest document conté el text Criteris, una explicació dels criteris que s'han seguit per elaborar el Corpus Oral Dialectal (COD). El COD és un component del Corpus de Català Contemporani de la Universitat de Barcelona (CCCUB), un arxiu de corpus de llengua catalana oral contemporània que ha estat confegit pel grup de recerca Grup d'Estudi de la Variació (GEV) amb la finalitat de contribuir a l'estudi de la variació dialectal, social i funcional en la llengua catalana.

Aquest i altres materials del CCCUB són accessibles directament al Dipòsit Digital de la UB (http://diposit.ub.edu) o a través del web del CCUB (http://www.ub.edu/cccub).

Aquesta obra està subjecta a la llicència de:
Reconeixement–NoComercial-SenseObraDerivada

http://creativecommons.org/licenses/by-nc-nd/3.0/es/deed.ca

LES DADES DEL QÜESTIONARI DEL COD

I. INTRODUCCIÓ GENERAL

El *Corpus Oral Dialectal* (COD) del català contemporani conté informació dels sis principals dialectes del català: alguerès, balear, central, nord-occidental, rossellonès i valencià. Les dades es van obtenir a partir d'una sèrie d'entrevistes que es van dur a terme a cada cap de comarca —o equivalent— del domini lingüístic català entre els anys 1994 i 1996. L'entrevista constava de dos blocs: la producció d'un text oral de tema lliure i la resposta a un qüestionari. Aquest qüestionari estava organitzat en diferents parts: aspectes fonètics rellevants, morfologia verbal regular, clítics pronominals, articles, possessius, pronoms personals forts, demostratius i locatius. Les dades obtingudes estan recollides en vuit bases de dades diferents (© Microsoft Access), cadascuna de les quals té una estructura adaptada al seu contingut. Les bases de dades que contenen tota la informació transcrita a partir de les respostes del qüestionari són les que presentem en aquest CD-ROM i estan emmagatzemades a la xarxa local del Departament de Filologia Catalana de la Universitat de Barcelona. Per a una descripció del mètode de recopilació de les dades de cada camp i de l'estructura de les bases de dades, vegeu els apartats corresponents.[1]

La informació recollida en les bases de dades correspon a un total de 303 entrevistes. Per regla general, es van entrevistar tres informants autòctons de cada localitat, l'edat dels quals estava compresa entre els 30 i els 45 anys. Aquesta franja d'edat assegurava que els enquestats no havien estat escolaritzats en català. A més, els informants tenien en comú el fet de pertànyer a la classe social mitjana, no haver passat llargues temporades fora de la localitat i no haver cursat estudis superiors. Per tal de preservar-ne l'anonimat, s'ha optat per identificar els informants mitjançant sigles formades per dues o tres grafies.

Les dades es transcriuen segons les convencions de l'Alfabet Fonètic Internacional (AFI). La transcripció és ampla: només s'usen diacrítics per tal de reflectir l'accent de mot. Tanmateix, en aquest cas, no s'empra el diacrític recomanat per l'AFI per transcriure l'accent ([ˈ]) sinó el que s'empra normalment per transcriure el to agut ([ˊ])

damunt la vocal de la síl·laba tònica. No s'accentuen els mots funcionals àtons, com ara les conjuncions *i*, *ni*, *si* i *que*; les preposicions *de*, *a*, *amb*; els articles, etc. Tampoc no es marquen els accents secundaris produïts per afebliment en una seqüència de mots, ni la desaccentuació en cas de xoc accentual (per tant, la seqüència *és més bo* es transcriu [éz méz βɔ́] i no [èz mez βɔ́], i la seqüència *va beure* es transcriu [bá βéwrə] i no [ba βéwrə]).

El nombre de registres que conté cada base de dades s'especifica a la taula següent:

Base de dades	Nombre de registres
Verbs	68262
Fonètica	55270
Clítics pronominals	24766
Articles	5049
Possessius	5223
Pronoms personals	1608
Demostratius	1512
Locatius	803
TOTAL	162.493

Ha estat interès del projecte de subministrar a l'usuari totes les dades obtingudes en cada resposta per cada localitat i, per això, es presenten, en general, tres jocs de dades per població, corresponents a les tres enquestes realitzades en cada localitat seleccionada. Per a cada pregunta de l'enquesta, doncs, les dades obtingudes poden ajustar-se a tres situacions diferents: coincidència total —situació ideal—, coincidència majoritària o divergència total. Però també ha estat interès del projecte de facilitar a l'usuari les dades simplificades i sistematitzades en el cas de les dades morfològiques, que contenen els paradigmes corresponents a la flexió verbal regular, els clítics pronominals, el articles, els possessius, els pronoms personals forts, els demostratius i els locatius. Per aconseguir aquest resultat, el projecte ha comptat amb un programa-filtre que ha seleccionat mecànicament les respostes majoritàries, seleccionades per identitats. Quan la divergència era total, s'ha triat manualment la resposta més

[1] La confecció del COD s'ha dut a terme en el marc dels projectes PB94-0951, PB97-0889, BFF2001-3798 i HUM2004-01504, tots ells finançacts pel MEC (i pel FEDER en el cas del darrer). Es pot consultar la informació sobre el projecte HUM2004-01504 a l'adreça <http://www.ub.edu/lincat>.

sistemàtica; és a dir, la que s'ajustava més al conjunt de les dades de la localitat. En el cas de les dades corresponents als aspectes fonètics, s'ha optat per presentar totes les dades recollides sense seleccionar-ne la forma majoritària, perquè aquestes no constitueixen paradigmes tancats sinó que recullen diverses realitzacions fonètiques dels segments en diferents contextos. A la pràctica, les bases de dades del CD-ROM contenen els registres obtinguts dels tres informants enquestats en cada localitat i, en les dades morfològiques, es ressalta, a més, la forma majoritària que ha resultat del filtratge —mecànic o manual— de totes les respostes. Les consultes es poden fer sobre el conjunt total de les dades i sobre les dades filtrades.

II. DESCRIPCIÓ DE LES BASES DE DADES DEL COD

II. 1 LA BASE DE DADES DE VERBS

La base de dades està constituïda pels cinc verbs model: *cantar* (1), *perdre* (2a), *beure* (2b), *sentir* (3a) i *servir* (3b). Aquesta classificació, que subdivideix els verbs de la segona i tercera conjugació en funció de l'absència / presència del segment extensió, es basa principalment en Viaplana (1984, 1986).[2] En principi, aquests verbs són els que conjuguen tots els informants; ara bé, si un informant té dificultats a l'hora de conjugar-ne algun, aquest se substitueix per un altre del mateix model (per exemple, *patir* per *servir*). El mètode utilitzat per tal d'obtenir les respostes ha estat la compleció d'oracions llegides per l'enquestador (*Cada dia jo ..... aigua*). Per norma general, la forma que l'informant diu espontàniament és la que es pren com a resposta vàlida i, per tant, la que s'introdueix a la base de dades.

A part dels cinc verbs model, en ocasions, s'han demanat verbs addicionals per confirmar respostes en què l'informant vacil·la (per exemple, davant del dubte entre [sentís] i [sentiɣés], es fa conjugar la forma corresponent de *dormir*) o bé per reflectir fenòmens característics del dialecte en qüestió (com l'harmonia vocàlica de [mɔ́ɣɔ] per

[2] Viaplana, Joaquim (1984) "La flexió verbal regular del valencià", dins Casanova, Emili (ed.), *Estudis en memòria del professor Manuel Sanchis Guarner: Estudis de Llengua i Literatura Catalanes*, I, p. 391-407, Universitat de València, València; Viaplana, Joaquim (1986) "Morfologia flexiva i flexió verbal catalana", *Llengua i Literatura* 1, p. 385-403. Altres treballs en la mateixa línia són: Lloret, Maria-Rosa i Joaquim Viaplana (1998) "Variació morfofonològica. Variants morfològiques", *Caplletra* 25, p. 43-62; Clua, Esteve (1998) *Variació i distància lingüística. Classificació dialectal del balear a partir de la morfologia flexiva*, tesi doctoral, Universitat de Barcelona; Perea, Maria Pilar (2000) *Estructura i*

moga). Quan l'informant conjuga tot o bona part del paradigma del verb addicional es crea una plantilla nova a la base de dades i el verb es classifica d'acord amb la conjugació i la presència o no d'extensió.[3] En valencià, els verbs addicionals que poden aparèixer conjugats, ja sigui de manera completa o parcial, són *moldre*, *coure*, *dormir*, *patir* i *moure*. Aquests tres últims també es poden trobar en nord-occidental, juntament amb *veure* i *plorar*. En el cas de l'alguerès, també es conjuguen els verbs *odiar*, *canviar*, *conèixer*, *dir*, *eixir* (usat com a verb model en lloc de *sentir*) i *dormir*. Com que en aquest últim dialecte els verbs conjugats són més nombrosos i el segment extensió és present en totes les conjugacions, s'ha adoptat la classificació següent: *cantar* (1a), *odiar* (1b), amb extensió palatal, i *canviar* (1c), amb extensió velar; *perdre* (2a), *conèixer* (2aa), sense extensió, i *beure* (2b), *dir* (2bb), amb extensió velar; *eixir* (3a), *dormir* (3aa), sense extensió, i *servir* (3b), amb extensió palatal.

La base de dades dels verbs consta de 19 caselles. Les tres primeres especifiquen el dialecte, la localitat i el codi de l'informant. A les quatre caselles següents s'indica la conjugació, la forma d'infinitiu, el temps i la persona de la forma verbal que es transcriurà en la casella FONÈTICA. A més, per tal de facilitar la cerca de les dades, s'inclou la forma ortogràfica corresponent a l'estàndard central.

Dialecte	Localitat	Inform	Conj	Verb	Temps	Persona	Ortografia
Central	Banyoles	CBF	1	cantar	Condicional	1	cantaria
Central	Banyoles	CBF	1	cantar	Condicional	2	cantaries
Central	Banyoles	CBF	1	cantar	Condicional	3	cantaria
Central	Banyoles	CBF	1	cantar	Condicional	4	cantaríem
Central	Banyoles	CBF	1	cantar	Condicional	5	cantaríeu
Central	Banyoles	CBF	1	cantar	Condicional	6	cantarien

Pel que fa a la casella FONÈTICA, cal tenir en compte que, tot i que l'informant normalment pronuncia les formes verbals dins d'una oració, aquestes es transcriuen aïlladament, sense reflectir, per exemple, les sonoritzacions que s'hagin pogut produir pel contacte entre el verb i el mot següent en la seqüència enregistrada. En el cas del pretèrit perfet perifràstic, sí que es reflecteixen els fenòmens d'assimilació ([vám bɔ̃wrə] per *van beure*, i [báŋ kantá] per *van cantar*) i de sonorització ([báz βɛ̃wrə] per *vas beure*) que pot originar el contacte entre l'auxiliar i l'infinitiu. Davant la fusió

variació en el verb català de començaments de segle: La flexió verbal en els dialectes catalans (Alcover-Moll), tesi doctoral, Universitat de Barcelona.

[3] Es crea la plantilla sempre que l'informant conjugui, com a mínim, el present d'indicatiu i de subjuntiu, l'imperfet d'indicatiu i de subjuntiu, i l'imperatiu. Si l'informant conjuga un nombre menor de temps verbals, aquests es transcriuran a la casella OBSERVACIONS.

de sibilants que sol afectar la segona persona del singular d'aquest mateix temps dels verbs *sentir* i *servir*, sempre s'ha optat per conservar sencera la forma del verb principal: [bá sərβí] per *vas servir*.

La forma transcrita a la casella FONÈTICA se segmenta morfològicament; el resultat d'aquesta segmentació es distribueix en les caselles següents: ARREL, EP (epèntesi immediatament posterior a l'arrel), EXT (extensió), TEMA (vocal temàtica), MODE (marca d'infinitiu, gerundi i participi), MT (mode/temps), EP2 (epèntesi posterior als morfemes de mode i mode/temps) i NP (nombre/persona). A continuació, s'inclou un fragment de la base de dades amb exemples de cadascuna d'aquestes caselles:

Fonetica	Arrel	Ep	Ext	Tema	Mode	MT	Ep2	NP
kəntəɾíə	kənt			ə		ɾíə		
pərðəɾém	pərð	ə				ɾé		m
bəvént	bə		v	é	nt			
báwɾə	báw				ɾ		ə	
sénten	sént						e	n
serβiʃko	serβ		iʃk			o		

Aquestes caselles, que recullen la forma fonètica i la segmentació morfològica, es deixen en blanc en alguns casos:

a) Quan l'informant no conjuga una forma o un temps verbal sencer, ja sigui perquè se n'oblida o perquè no li surt.
b) Quan l'informant s'equivoca de temps o de persona.
c) Quan l'informant va massa ràpid i no acaba de conjugar un verb.

En totes aquestes situacions es fa el comentari pertinent a la casella OBSERVACIONS. A més, aquesta casella pot incloure comentaris addicionals que susciten les formes verbals elicitades per l'informant: algun aspecte fonètic o fonològic destacable, apreciacions dels informants amb relació a l'ús (factors generacionals, socials o geogràfics) i comentaris referents als dubtes i dificultats que pot tenir l'informant a l'hora de conjugar una forma. També s'hi col·loquen les formes de determinats temps verbals que l'enquestador demana com a informació complementària (vegeu nota 2).

A continuació, es recullen alguns dels enunciats més habituals de la casella OBSERVACIONS:

- L'informant no conjuga aquesta persona.
- L'informant es confon: diu [x].

- L'informant va massa ràpid i no acaba la forma.
- L'informant utilitza indistintament les dues formes.
- L'informant comenta que utilitza més la forma [x].
- L'informant també accepta la forma proposada per l'enquestadora: [x].
- A l'informant també se li demana que conjugui el verb 'dormir': [x].
- L'informant explica que la forma pròpia de la zona és [x].

Finalment, cal fer referència a la casella ALTERNATIVES, on es transcriuen les formes alternatives proposades per l'informant —si n'hi ha més d'una apareixen separades per comes. Les alternatives solen ser:

a) Formes que l'informant pot utilitzar indistintament amb les elicitades de manera espontània.
b) Formes que l'informant utilitza amb menys freqüència.
c) Variants més estandarditzades que usa en un context formal.
d) Formes marcades des d'un punt de vista generacional.
e) Formes que són pròpies de la localitat de l'informant però que aquest ha deixat de fer servir, generalment perquè s'ha autocorregit.

Tal com ja s'ha dit, si els informants fan comentaris amb relació a l'ús d'aquestes formes, es recullen a la casella OBSERVACIONS. Convé remarcar que a la casella ALTERNATIVES només s'han transcrit aquelles formes que l'informant ha pronunciat. En els casos en què l'informant accepta formes alternatives proposades per l'enquestador, però no les arriba a pronunciar, es deixa en blanc la casella ALTERNATIVES però se'n deixa constància a les OBSERVACIONS.

II. 2 LA BASE DE DADES DE FONÈTICA

Aquesta base de dades recull la transcripció fonètica d'una sèrie de mots aïllats i de seqüències de mots. A partir d'aquestes dades, s'han volgut recollir les diverses realitzacions fonètiques que pot tenir un o més segments en un context determinat, en els sis principals dialectes del català. En el cas de les seqüències de mots, l'objecte específic d'estudi són els processos fonològics que poden actuar entre mots, com ara les

sonoritzacions i les assimilacions. En general, el mètode utilitzat per obtenir les dades era la presentació d'una sèrie de dibuixos o fotografies que l'informant havia d'identificar. Quan això no era possible es treballava en la compleció d'oracions, per exemple, *El contrari de tou és* .... Finalment, pel que fa a les seqüències de mots, s'utilitzava un altre procediment: l'informant havia de traduir al català una seqüència de mots redactada en castellà (en el cas del català central, nord-occidental, valencià i balear), en francès (en el cas del rossellonès) o en italià (en el cas de l'alguerès).

El balear, l'alguerès i el rossellonès compten amb un nombre més elevat de registres que la resta de dialectes. Amb aquesta informació addicional s'han volgut recollir fenòmens característics d'aquests dialectes.

La base de dades de fonètica està encapçalada per les caselles —comunes a totes les bases de dades— INFORMANT, LOCALITAT i ID. Aquestes caselles recullen, respectivament, el codi de l'informant, la localitat a la qual pertany i l'identificador numèric que correspon a cada registre. La resta de caselles de la base de dades de fonètica es presenten a continuació:

MOT_ANT: conté la transcripció ortogràfica dels mots que precedeixen el mot o la seqüència de mots objecte d'estudi.

MOT_ANT_FON: conté la transcripció fonètica dels mots que precedeixen el mot o la seqüència de mots objecte d'estudi.

ARREL: conté la transcripció fonètica de l'arrel del mot objecte d'estudi, p.ex. [**ruð**étə]. [4]

VOC_INTER: conté la transcripció fonètica de la vocal (si n'hi ha cap) que precedeix la nasal final d'algunes arrels, p. ex. [ɔmənːs].

NASAL: conté la transcripció fonètica de la consonant final nasal (si n'hi ha cap) d'algunes arrels, p. ex. [ruðónə].

DIMINUT: conté la transcripció fonètica del morf de diminutiu (si n'hi ha cap), p. ex. [ʎiβɾ**ét**].

VOC_FINAL: conté la transcripció fonètica de la vocal final àtona (si n'hi ha cap) que no forma part de l'arrel, p. ex. [kázə], [rəʎɔ́dʒə].

PLURAL: conté la transcripció fonètica del morf de plural (si n'hi ha cap).

[4] Els mots que presenten una *–n* a l'arrel (tipus *hòmens*) es tracten de manera diferent. Vegeu les caselles següents.

SEQÜÈNCIA: conté la transcripció fonètica de la seqüència de mots en qüestió (vegeu nota 1), p.ex. [péjʒ βláw]. Aquesta casella està destinada únicament a les seqüències de mots, les quals es presenten transcrites fonèticament. En aquests casos, les caselles anteriors restaran buides.

SENCER: conté la transcripció fonètica del mot complet no segmentat morfològicament, p.ex. [ʎiβɾét].

FORMES_ALT.: conté la transcripció fonètica de les formes alternatives proposades per l'informant.

OBSERVACIONS: conté dades de caràcter bastant divers. S'hi transcriuen les respostes que no aporten la informació corresponent a la pregunta formulada o comentaris rellevants que afegeixen informació o matisen la resposta de l'informant. També s'hi inclouen comentaris referents a la vacil·lació de l'informant entre dues o més formes, a la dificultat de comprensió d'alguna reposta, etc. A continuació, es presenta una mostra dels continguts més habituals de la casella OBSERVACIONS:

- L'informant diu [x].

  La resposta es transcriu íntegrament i exclusivament en aquesta casella; es tracta dels casos en què la resposta de l'informant no és vàlida, atès que en el mot que l'informant pronuncia no hi ha el so esperat o el context en el qual interessa observar aquest so.
- En primer lloc l'informant diu la forma següent, que rectifica espontàniament: [x].

  En aquests casos l'informant diu una forma en primer lloc, però la rectifica espontàniament per una altra, que es transcriu a les caselles principals. La forma pronunciada en primer lloc es transcriu a les OBSERVACIONS.
- La resposta no s'entén.
- Aquesta pregunta no es formula a l'informant.
- L'informant diu _____ en lloc de _____. P.ex., *L'informant diu el femení en lloc del masculí*.

  Aquest tipus d'observacions es fan servir en els casos en què, to i que l'informant canvia el gènere o el nombre (en el cas dels noms i adjectius) o el temps o la persona (en el cas dels verbs), el so o el contacte de sons esperats es mantenen. En aquests casos la forma que pronuncia l'informant es transcriu íntegrament a les caselles principals.
- La consonant final se sonoritza per contacte amb la vocal / consonant sonora següent.

En els casos en què el mot esperat acaba amb una consonant que esdevé pel contacte amb la vocal o la consonant sonora inicial del mot que l'informant pronuncia a continuació. S'ha optat per transcriure la consonant final com a sorda. És el cas de: *dit**s*** [ďits] + *de la mà*, *camp**s*** [káms] + *de futbol*, etc., en què els mots objecte d'estudi són únicament *dits* i *camps*.[5]

En el cas de les seqüències de mots en les quals l'informant canvia el segon mot per un altre que manté les mateixes característiques fonètiques que el proposat per l'enquestador, la resposta no es transcriu a les observacions, sinó a les caselles principals. Per exemple: *serp intel·ligent* per *serp hàbil*, en les quals el context que es vol estudiar és *–rp + vocal*.

Pel que fa al sistema de cerca de dades, es disposa d'un conjunt de caselles que recullen la informació morfològica del mot objecte d'estudi i el tret fonètic que interessa observar en cada cas. Aquestes caselles estan marcades amb els valors 1 (presència de tret) o 0 (absència de tret). A continuació se'n presenta la llista:

a) Informació morfològica:

- SUBST_S: substantiu singular
- SUBST_M: substantiu masculí
- SUBST_F: substantiu femení
- SUBST_PL: substantí plural
- DIM: diminutiu
- ADJ_S: adjectiu singular
- ADJ_M: adjectiu masculí
- ADJ_F: adjectiu femení
- ADJ_PL: adjectiu plural
- VERB
- ALTRES: en els casos en què el mot no pertany a cap de les categories anteriors.
- CONTACT: en els casos de seqüències de mots

b) Tret fonètic

- VOC_INICIAL: vocal inicial
- VOC_PRE: vocal pretònica

[5] Aquesta observació només és vàlida per als mtos aïllats, i no per a les seqüències de mots en què precisament interessen els processos fonològics produïts per contacte, p.ex. [pɛ́z ənórmə].

- E_TON: *e* tònica
- O_TON: *o* tònica
- VOC_AL_TON: qualsevol vocal tònica, excepte *e* i *o*.
- VOC_POST: vocal posttònica
- VOC_FIN: vocal final
- GLIDES
- CONS. OCLUSIVA / APROXIMANT
- CONS. FRICATIVA
- CONS. AFRICADA
- CONS. NASAL
- CONS. LATERAL
- CONS. RÒTICA
- GR_CONS_IN: grup consonàntic inicial
- GR_CONS_ME: grup consonàntic medial
- GR_CONS_FI: grup consonàntic final
- GRUP_HETE: grup heterorgànic
- GRUP_HOMO: grup homorgànic

II. 3 LA BASE DE DADES DE CLÍTICS PRONOMINALS

La base de dades de clítics pronominals conté informació provinent d'un qüestionari que consta d'una sèrie d'oracions redactades en castellà —en el cas del Principat, del País Valencià, d'Andorra i de les Illes Balears—, en francès —en el cas del rossellonès— i en italià —a l'Alguer. A partir de la traducció al català d'aquestes oracions, s'esperava aconseguir l'aparició d'un pronom o d'una seqüència de pronoms en un context determinat. D'aquesta manera, s'han obtingut les formes que adopten els pronoms febles *em*, *et*, *es*, *el*, *la*, *els* —acusatiu de tercera persona del plural masculí —, *les*, *ens*, *us*, *en*, *hi* i *ho* en procilisi prevocàlica i preconsonàntica i en enclisi postvocàlica i postconsonàntica. També s'ha obtingut informació sobre les combinacions de datius i acusatius de tercera persona, tant en posició proclítica (sense tenir en compte si el verb comença en vocal o en consonant) com enclítica (postvocàlica i postconsonàntica). El nombre d'oracions presentades varia segons el dialecte.

A la base de dades s'ha assignat un identificador numèric a cada oració (ID_PF). A continuació, s'han segmentat les respostes en les caselles següents, presentades segons l'ordre lineal —d'esquerra a dreta— en què apareixen:

MOTS_ANT: conté la transcripció ortogràfica de la seqüència de mots que precedeix el grup format pel clític i el verb.

MOTS_ANT_FON: conté la transcripció fonètica de la seqüència de mots que precedeix el grup format pel clític i el verb.

V_I1: conté la transcripció fonètica de la vocal (si n'hi ha cap) que precedeix la consonant del primer pronom, p. ex. [kwán əd réntu stáz miʎó].

PRO1: conté la transcripció fonètica de la consonant del primer pronom o de la vocal dels pronoms *hi* i *ho*, p. ex. [kwán əd réntu stáz miʎó] i [ə réwz j áz ð əná suβín].[6]

V_FIPRE1: conté la transcripció fonètica de la vocal (si n'hi ha cap) que segueix la consonant del primer pronom, p. Ex. [lə səlúðə káðə ðíə].

PLUR1: conté la transcripció fonètica de la consonant del morf de plural (si n'hi ha cap) del primer pronom feble, p. ex. [ləz ɔ́βrə].

V_FIPOST1: conté la transcripció fonètica de la vocal (si n'hi ha cap) que segueix el morf de plural del primer pronom, p. ex. [əlzə méɲʒə].

V_I2: conté la transcripció fonètica de la vocal (si n'hi ha cap) que precedeix la consonant del segon pronom, p. ex. [əlz əŋ kumpɾəɾé wn tɾɔ́s].

PRO2: conté la transcripció fonètica de la consonant del segon pronom o de la vocal dels pronoms *hi* i *ho*, p. ex. [əlz əŋ kumpɾəɾé wn tɾɔ́s] i [əlz i kómpɾəs].

V_FIPRE2: conté la transcripció fonètica de la vocal (si n'hi ha cap) que segueix la consonant del segon pronom, p. ex. [li les kompɾaɾé].

PLUR2: conté la transcripció fonètica de la consonant del morf de plural (si n'hi ha cap) del segon pronom, p. ex. [li les kompɾaɾé].

V_FIPOST2: conté la transcripció fonètica de la vocal (si n'hi ha cap) que segueix el morf de plural del segon pronom, p. ex. [s əlzə méɲʒə kaðə ðíə].

[6] En aquesta casella també es transcriu el morf de plural quan s'anteposa a la consonant del pronom de tercera persona (cosa que succeeix en algunes varietats del menorquí). Per exemple, [əzð ɔ́bɾə] per *els obre*.

VOC_VERB: conté la transcripció fonètica del verb començat en vocal (només en els casos de proclisi), p. ex. [m **əréɣlu**]. També conté la transcripció fonètica del verb començat en una vocal que s'ha elidit, p. ex. [li **splikəré**].

CONS_VERB: conté la transcripció fonètica del verb començat en consonant (només en els casos de proclisi), p. ex. [əm **réntu**].

VERB_VOC: conté la transcripció fonètica del verb acabat en vocal (només en els casos d'enclisi), p. ex. [**réntə**m]. També conté la transcripcó fonètica del verb acabat en una vocal que s'ha elidit, p. ex. [**pɔ́z**u].

VERB_CONS: conté la transcripció fonètica del verb acabat en una consonant que no sigui *–nt*, *–r* o *–u* semivocal (només en els casos d'enclisi), p. ex. [**féz**nə].

VERB_W: conté la transcripció fonètica del verb acabat en *–u* semivocal (només en els casos d'enclisi), p. ex. [**ərəɣlɛ́w**zə].

VERB_R: conté la transcripció fonètica del verb acabat en *–r* (només en els casos d'enclisi), p. ex. [**rəntár**mə]. També conté la transcripció fonètica d'un verb que acaba en una *r* que s'ha elidit, p. ex. [**rəntá**lzə].

VERB_NT: conté la transcripció fonètica del verb acabat en *–nt* (només en els casos d'enclisi), p. ex. [**pləɲʃan**lés].

VO_I1: conté la transcripció fonètica de la vocal (si n'hi ha cap) que precedeix la consonant del primer pronom en els casos d'enclisi, p. ex. [dunéwəlzi].

P_1: conté la transcripció fonètica de la consonant del primer pronom (si n'hi ha cap) en els casos d'enclisi o de la vocal dels pronoms *hi* i *ho*, p. ex. [dunéwə**l** zi] i [ərîβə **j** əβiát].

VO_FE1: conté la transcripció fonètica de la vocal (si n'hi ha cap) que segueix la consonant del primer pronom en els casos d'enclisi, p. ex. [əkumpáɲələ].

PL1: conté la transcripció fonètica de la consonant del morf de plural (si n'hi ha cap) del primer pronom en els casos d'enclisi, p. ex. [dunéwəl**z**i].

VO_FST1: conté la transcripció fonètica de la vocal (si n'hi ha cap) que segueix el morf de plural en els casos d'enclisi, p. ex. [dunɛ́wəlz**ə**].

VO_I2: conté la transcripció fonètica de la vocal (si n'hi ha cap) que precedeix la consonant del segon pronom en els casos d'enclisi, p. ex. [əmbíənəlzi].

P_2: conté la transcripció fonètica de la consonant del segon pronom (si n'hi ha cap) en els casos d'enclisi o de la vocal dels pronoms *hi* i *ho*, p. ex. [dónali**l** es] i [kompɾálz**i**].

VO_FE2: conté la transcripció fonètica de la vocal (si n'hi ha cap) que segueix la consonant del segon pronom en els casos d'enclisi, p. ex. [dónalil**e**s].

PL2: conté la transcripció fonètica de la consonant del morf de plural (si n'hi ha cap) del segon pronom en els casos d'enclisi, p. ex. [dónalile**s**].

VO_ST2: conté la transcripció fonètica de la vocal (si n'hi ha cap) que segueix el morf de plural en els casos d'enclisi, p. ex. [əmbíənəlz**i**].

MOTS_POST: conté la transcripció ortogràfica de la seqüència de mots que segueix el grup format pel clític i el verb.

MOTS_POST_FON.: conté la transcripció fonètica de la seqüència de mots que segueix el grup format pel clític i el verb.

SENCER: conté la transcripció fonètica del pronom feble (només en els casos de proclisi).

OBSERVACIONS: conté dades de caràcter bastant divers. S'hi transcriuen des de les respostes que no aporten la informació corresponent a la pregunta formulada (vegeu més avall) fins a les formes alternatives aportades per l'informant. A més a més, s'hi inclouen alguns comentaris com ara la vacil·lació de l'informant entre dues o més formes, la dificultat de comprensió d'alguna resposta, etc. A continuació, es presenta una mostra dels continguts més habituals d'aquesta casella:

- L'informant també diu la forma [x].
- En primer lloc l'informant diu la forma següent, que rectifica espontàniament: [x].
- L'informant vacil·la entre aquesta forma i la forma [x].
- L'informant utilitza indistintament aquesta forma i la forma [x].
- L'informant explica que entre la gent gran se sent la forma [x].
- La resposta no s'entén gaire bé.

Sempre que la resposta de l'informant no es correspon amb l'esperada, s'ha optat per no segmentar la seqüència i s'ha transcrit íntegrament a la casella OBSERVACIONS. Els casos concrets són els següents:

a) Quan, d'acord amb la pregunta, s'espera que els pronoms apareguin en posició enclítica i, en canvi, apareixen en posició proclítica o a la inversa (per exemple, la resposta *Les va haver d'acompanyar al col·legi* a una oració com *Tuvo que acompañarlas al colegio*).
b) Quan, d'acord amb la pregunta, s'observa que l'informant canvia el gènere o el nombre del sintagma que actua de referent del pronom (és el cas d'una resposta com *Aquesta nina, no t'oblidis de comprar-la a les nenes* a l'oració *Este muñeco, no te*

olvides de comprárselo a los niños — que indica que s'espera informació referent al pronom d'acusatiu de tercera persona del singular masculí i no pas femení).

c) Quan, d'acord amb la pregunta, s'espera que els pronoms proclítics apareguin en posició prevocàlica i, en canvi, apareixen en posició preconsonàntica, o a la inversa (per exemple, *Sempre em vesteixo quan surto de casa* com a resposta a la pregunta *Siempre me arreglo al salir de casa*).
d) Quan, en els casos d'enclisi, els pronoms apareixen posposats a una forma verbal diferent de la que s'espera segons la pregunta (per exemple, *Renteu-me el cotxe* com a resposta a la pregunta *Lávame el coche*).

II.3.1 Criteris de transcripció ortogràfica

En la transcripció ortogràfica, pel que fa a la morfologia, es mantenen les realitzacions dialectals de la flexió verbal i nominal (per exemple, *cridos* per *cridis*, *faigues* per *facis*, etc.). També es mantenen les formes dialectals de l'article definit masculí *lo*, *los* i les diverses formes de l'article salat. Els pronoms febles que precedeixen un verb començat en consonant es transcriuen ortogràficament amb la forma reforçada. No es reflecteixen determinades pronúncies de seqüències de verb i clític com ara *gità's* o *vendre-lo*, que es transcriuen ortogràficament com *gitar-se* i *vendre'l*.

Pel que fa als aspectes lèxics, es transcriuen els dialectalismes (per exemple, *sàller* i *sallir* del rossellonès per *sortir*) i les interferències d'altres llengües. En aquest últim cas, s'ha optat per transcriure les formes provinents del francès o de l'italià d'acord amb l'ortografia d'aquestes llengües (per exemple, *voiture* per *cotxe* o *assai* per *molt*); en canvi, les formes d'origen espanyol s'escriuen amb l'ortografia adaptada a les convencions del català (per exemple, *unya* per *ungla*). Les variants formals (com ara *cotxo*, *aigo-aiga-auia*, *lego* per *luego*, *aülla*, *llavò-llavòs-llavons-llavòrens-llavonses*) es transcriuen amb una sola forma, la que apareix al DIEC, per tal de facilitar al màxim les cerques.

II.3.2 Criteris de transcripció fonètica

La seqüència fònica es divideix en una sèrie de caselles independents. Quan hi ha elisions o fusions d'elements que pertanyen a dues caselles diferents, s'ha hagut d'optar

per assignar l'únic element que es percep a una de les dues caselles. A continuació, es detallen els criteris que s'han aplicat en aquests casos.

Sempre s'ha prioritzat la seqüència formada pel verb i el clític. Així, en una seqüència com [káðə ði **ə**m réntu lə káɾə] (*Cada dia em rento la cara*), la segona vocal neutra, tot i que pot pertànyer tant al substantiu *dia* com al pronom feble de primera persona, ha estat assignada a aquest últim. S'aplica el mateix criteri quan hi ha coincidència de la vocal final del pronom enclític amb la vocal inicial de la seqüència de mots posteriors. En una seqüència com *la bicicleta* [kómpɾalil**a** le tʃikétes] (*compra-li-la a les xiquetes*), la segona *a* és transcrita a la casella que correspon a la vocal final del pronom i no a la dels mots posteriors.[7]

Com a conseqüència de la segmentació de la cadena fònica en diferents caselles, es pot trobar alguna cel·la en què el mot inicial comença en una seqüència de sons que no forma una estructura sil·làbica permesa en català. És el cas de la seqüència [kwán ed rénto **st**áz miʎó] (*Quan et rento estàs millor*). També és possible trobar el cas invers, és a dir, que la seqüència d'una casella acabi en una estructura sil·làbica no permesa en català, per exemple, [sém**pr** əz réntə ləz ðéns] (*Sempre es renta les dents*).

En les caselles de mots anteriors i mots posteriors, la cadena fònica es presenta seqüenciada en mots lèxics, cadascun dels quals es transcriu amb un accent primari. En alguns casos, s'han adoptat certs criteris ortogràfics; així, els proclítics s'han transcrit separadament del verb que acompanyen (per exemple, [əm réntu] per *em rento*). En canvi, els pronoms enclítics s'han transcrit juntament amb el verb que acompanyen (per exemple, [réntəm] per *renta'm*).

D'altra banda, dins una mateixa casella es poden trobar contactes de vocals idèntiques que es resolen amb l'elisió d'una de les vocals. Quan això succeeix, si una de les dues vocals és lèxica i l'altra és epentètica, es transcriu la lèxica. Quan les dues vocals són lèxiques, es poden donar diverses possibilitats. Si una és tònica i l'altra és àtona, es transcriu la tònica, per exemple, [akét ɔ́m **é**z el métʃe] (*Aquest home és el metge?*). Quan les dues vocals són àtones, es prioritza la integritat del mot situat més a la dreta (per

[7] Aquest criteri s'ha seguit a totes les bases de dades que contenen la transcripció fonètica dels mots anteriors i/o posteriors al mot objecte d'estudi.

exemple, [m aɣɾáð **a**riβáɾ] per *m'agrada arribar*), excepte en el cas que només el mot de l'esquerra pertanyi a una de les categories lèxiques majors (per exemple, [kómpɾə m təɾʒétə] per *compra amb targeta*).

En el cas de contacte de consonants, sempre es transcriu la consonant que es troba en posició d'obertura sil·làbica (per exemple, [mé **s**uβín] per *més sovint*; [reunjón **s**émpɾe] per *reunions sempre*).

II. 4 LA BASE DE DADES D'ARTICLES

Aquesta base de dades consta de 24 registres corresponents a les formes de l'article determinat, l'article personal i l'article determinat masculí precedit per les preposicions *per*, *amb*, *a* i *de*. Els 16 primers registres, que recullen les formes de l'article determinat i de l'article personal, són comuns a tots els informants; els 8 registres restants, que corresponen a les formes de l'article determinat masculí precedit per les preposicions *per*, *amb*, *a* i *de*, només es recullen en el cas de l'alguerès i el balear.

Pel que fa als contextos d'aparició, tant l'article determinat com l'article personal s'han recollit seguits de mot començat en vocal i en consonant. En el cas de l'article determinat femení singular també s'han tingut en compte, específicament, els contextos en què precedeix un mot començat en les vocals *i* o *u* àtones i *i* o *u* tòniques.

Les respostes es van obtenir mitjançant la traducció al català de seqüències que contenien els articles en castellà —per al balear, el central, el nord-occidental i el valencià—, en francès —per al rossellonès— i en italià —per a l'alguerès.
A part de les caselles comunes a totes les bases de dades (INFORM, LOCAL i ID), aquesta s'organitza en les caselles següents:
MOT_ANT: conté la transcripció fonètica de la seqüència de mots (si n'hi ha cap) que precedeix l'article o la contracció preposició + article.
CONS_ANT: conté la transcripció fonètica de la consonant inicial de les preposicions *per* i *de*, quan formen contracció amb l'article determinat masculí singular o plural.
VOC_ANT: conté la transcripció fonètica de la vocal (si n'hi ha cap) que precedeix la consonant de l'article.

ART: conté la transcripció fonètica de la consonant de l'article.[8]

VOC_FINAL: conté la transcripció fonètica de la vocal (si n'hi ha cap) que segueix la consonant de l'article.

PLURAL: conté la transcripció fonètica de la consonant del morf de plural (si n'hi ha cap) de l'article. En el cas de l'article masculí plural *es*, s'ha optat per transcriure la fricativa final a la casella ART i no pas a la casella de PLURAL.

MOT_POST: conté la transcripció fonètica de la seqüència de mots (si n'hi ha cap) que hi ha a continuació de l'article.

SENCER: conté la transcripció fonètica de l'article o de la contracció preposició + article.

OBSERVACIONS: conté dades de caràcter bastant divers. S'hi transcriuen respostes que no aporten la informació corresponent a la pregunta formulada, alternatives proposades per l'informant, comentaris referents a la vacil·lació entre dues formes, a errors comesos per l'informant, etc.

Per tal de facilitar el sistema de cerques, la base de dades conté una casella amb la forma ortogràfica corresponent a l'estàndard central de l'article determinat, l'article personal o l'article determinat masculí precedit per les preposicions *per*, *amb*, *a* i *de*, i del mot que els segueix. També disposa d'una sèrie de caselles amb informació morfològica referent al gènere i al nombre (MASC, FEM, SING i PL).

II. 5 LA BASE DE DADES DE POSSESSIUS

Aquesta base de dades recull les diverses formes que adopten els possessius quan s'usen anteposats al nom que determinen. Es recullen tant les formes dels possessius tònics —d'un sol posseïdor i de diversos posseïdors— com les dels possessius àtons, tot i que aquestes últimes no es pregunten sistemàticament a tots els informants. D'altra banda, els possessius de 3a persona que indiquen més d'un posseïdor *llur*, *llura*, *llurs* i *llures* tan sols es pregunten al dialecte rossellonès. En tots els casos, el mètode usat per obtenir les dades va ser la traducció d'una sèrie de seqüències breus que contenien les formes corresponents dels possessius.

A part de les caselles comunes a totes les bases de dades (INFORM, LOCALITAT i ID), aquesta s'organitza en les caselles següents:

[8] En aquesta casella, en el cas de Pollença és possible trobar-hi la vocal [u], característica de l'article d'aquesta varietat.

VOC_ANT: conté la transcripció fonètica de la vocal (si n'hi ha cap) que precedeix la consonant de l'article que apareix anteposat al possessiu.
ART: conté la transcripció fonètica de la consonant de l'article.
VOC_FIN1: conté la transcripció fonètica de la vocal (si n'hi ha cap) que segueix la consonant de l'article.
PLURAL1: conté la transcripció fonètica de la consonant del morf de plural (si n'hi ha cap) de l'article.
POSS: conté la transcripció fonètica de l'arrel del possessiu.
VOC_FIN2: conté la transcripció fonètica de la vocal corresponent a les epèntesis o als morfs de flexió (si n'hi ha cap) del possessiu.
PLURAL2: conté la transcripció fonètica de la consonant del morf de plural (si n'hi ha cap) del possessiu.
MOT_POST: conté la transcripció fonètica del substantiu que es pronuncia després del possessiu.
SENCER: conté la transcripció fonètica del possessiu.
OBSERVACIONS: conté informació diversa, com ara formes alternatives proposades pels informants, respostes no esperades o comentaris referents a l'ús d'una forma.

Per tal de facilitar el sistema de cerques, la base de dades conté una casella amb la forma ortogràfica corresponent a l'estàndard central del possessiu i de l'article que el precedeix. També disposa d'una sèrie de caselles que indiquen el gènere i el nombre del possessiu (MASC, FEM, SING i PLUR).

II.6 LA BASE DE DADES DE PRONOMS PERSONALS

Aquesta base de dades conté vuit registres per informant, que corresponen als pronoms personals singulars i plurals de primera, segona i tercera persona (masculí i femení). Les dades es van obtenir a partir de la traducció al català d'una sèrie de seqüències breus que contenien les formes corresponents dels pronoms personals.

La base de dades de pronoms personals està encapçalada per les caselles comunes a totes les bases de dades (INFORM, LOCAL i ID). La resta de caselles es presenten a continuació:
T_FON: conté la transcripció fonètica del pronom personal.

MOT_POST: conté la transcripció fonètica de la seqüència de mots que segueix el pronom personal.
OBSERVACIONS: recull informació de caràcter divers, com ara les formes alternatives proposades per l'informant o les formes que no corresponen a la resposta esperada, entre d'altres.

Per tal de facilitar el sistema de cerques, aquesta base de dades conté una casella amb la forma ortogràfica del pronom personal corresponent a l'estàndard central. També disposa d'una sèrie de caselles amb informació morfològica referent al gènere, al nombre i a la persona (SING, MASC, FEM, PL, PERS).

II. 7 LA BASE DE DADES DE DEMOSTRATIUS NEUTRES

Aquesta base de dades conté tres registres per informant, que corresponen als tres possibles graus de proximitat dels demostratius neutres, representats per les formes *açò* (1r grau), *això* (2n grau) i *allò* (3r grau). Les dades es van obtenir a partir de la traducció al català d'una sèrie de seqüències breus que contenien les formes corresponents dels demostratius neutres.

A part de les caselles comunes a totes les bases de dades (INFORM, LOCAL i ID), aquesta disposa de les caselles següents:
MOT_ANTERIOR: conté la transcripció fonètica del mot anterior al demostratiu.
FONÈTICA: conté la transcripció fonètica del demostratiu.
OBSERVACIONS: s'hi recullen les formes alternatives o les formes que no corresponen a la resposta esperada, entre d'altres.

Per tal de facilitar el sistema de cerques, aquesta base de dades compta amb una casella que indica el grau de proximitat del demostratiu neutre (1, 2 o 3).

II. 8 LA BASE DE DADES DE LOCATIUS

Aquesta base de dades conté tres registres per informant, que corresponen als tres possibles graus de proximitat dels locatius, representats per les formes *ací* (1r grau),

aquí (2n grau) i *allí* / *allà* (3r grau). Les dades es van obtenir a partir de la traducció al català d'una sèrie de seqüències breus que contenien les formes corresponents dels locatius.

A part de les caselles comunes a totes les bases de dades (INFORM, LOCAL i ID), aquesta disposa de les caselles següents:

MOT_ANTERIOR: conté la transcripció fonètica del mot que precedeix el locatiu.

TRANS_FON: conté la transcripció fonètica del locatiu.

OBSERVACIONS: recull informació de caràcter divers, com ara les formes alternatives que proposa l'informant o les formes que no corresponen a la resposta esperada, entre d'altres.

Pel que fa al sistema de cerques, aquesta base de dades compta amb una casella que indica el grau de proximitat del locatiu (1, 2 o 3).