

Criteris per a la transcripció del corpus *Parlars*

1. Introducció

La transcripció “és un procediment de trasllat o transposició a una forma escrita d’unes dades que originalment s’han produït a través del canal oral” (Payrató 2010: 208). Hi ha diversos tipus de transcripció: la transcripció pròpiament dita i la transliteració ortogràfica (vg. Hidalgo & Sanmartín 2005, Bladas 2009). La transcripció és un model més detallat que pretén reflectir la llengua oral sense perdre gaire informació o el mínim possible, com ara la transcripció fonètica amb AFI o altres models amb símbols prosòdics sobre pauses, encavalcaments, gestualitats entre altres aspectes vocals (vg. COC, Alturo & Payrató 2002; Val.Es.Co, Briz 2002; Bladas 2009). La transliteració, en canvi, és un model que usa les normes ortogràfiques convencionals i, per tant, perd part de la informació com ara elements prosòdics (vg. la transcripció fonoortogràfica del COD, Viaplana & Perea 2003; *l’Atles entonatiu*, Prieto & Cabré 2007-2012; el corpus COSER, de Benito, Pueyo & Fernández-Ordóñez 2016; o el MC-NLCH, Samper, Hernández & Troya 1998).

Els criteris de transcripció que presentem s’ajusten als objectius del corpus *Parlars* (projecte CorDiVal: *Elaboració d’un corpus oral dialectal del valencià col·loquial*, Ref. GV/2017/094, finançat per la Generalitat Valenciana; <https://www.uv.es/corvalc/>), és a dir, un corpus orientat a l’estudi de la variació dialectal i diafàsica de la llengua oral. També s’han tingut en compte els criteris d’altres projectes anteriors, especialment els del *Corpus de Català Contemporani de la Universitat de Barcelona* (Payrató & Alturo 2002, Viaplana et al. 2003, Viaplana & Perea 2003, Foix-Fuster et al. 2007, Alturo et al. 2009, Carrera-Sabaté & Viaplana, Perea & Viaplana, Pons & Viaplana 2009, Blesa 2009, Payrató 2010) i els de *l’Atles interactiu de l’entonació del català* (Prieto & Cabré 2007-2012). Finalment, també ens hem basat en les característiques lingüístiques dialectals generals (Veny 1982, Solà et al. 2002, Beltran & Segura 2017).

El corpus *Parlars* tindrà dues transcripcions: una *transcripció fonoortogràfica* que provarà de reflectir al màxim possible les característiques lingüístiques del discurs i una *transcripció ortogràfica estandarditzada* que ha de facilitar les tasques d’anotació. Aquest model serà possible perquè optem per una estratègia d’anotació multicapes (*stand-off*) que combina la superposició de diversos nivells de transcripció i anotació alineats amb el document audiovisual original. La font primària del corpus és l’arxiu de so (o vídeo) dividit en segments breus que faciliten la cerca d’elements lingüístics transcrits i el fragment corresponent del document multimèdia on hi apareixen. La transcripció es realitza amb l’eina ELAN (Brugman & Russel 2014) (vg. les instruccions del funcionament del programa amb els dos tutorials que hem preparat per a l’elaboració del corpus *Parlars*, Esplà et al. 2018 i Esplà et al. 2019). L’estratègia d’anotació multicapa permet afegir diverses capes de transcripció o anotació segons les necessitats. Les diverses capes que s’introdueixen dependran de cada text, però previsiblement hi ha almenys les següents:

- Una capa amb els silencis més llargs d’un determinat llindar (vg. la primera part del tutorial, Esplà et al. 2018).
- Una capa amb la transcripció fonoortogràfica per cada parlant.
- Una capa amb la transcripció ortogràfica estandarditzada per cada parlant. Serà una línia secundària (vg. la segona part del tutorial, Esplà et al. 2019).
- Una capa amb la segmentació per tokens i la categoria gramatical, basada en la transcripció ortogràfica. La lematització es durà a terme amb l’anàlitzador morfològic proporcionat per la col·lecció d’eines Apertium (Forcada et al. 2011).

En la mesura que avança el projecte, podrem afegir-hi noves capes amb anotació de fenòmens diversos (fonètica i prosòdia, sintaxi, semàntica, pragmàtica, etc.) (Nivre 2008; Nissim & Pietrandrea 2017).

En definitiva, el model de transcripció que presentem no tindrà un fitxer en format .doc o .txt, és a dir, no serà un text lineal de l'àudio en forma de text escrit. Per tant, no hi haurà trets tipogràfics (negreta, cursiva, subratllat, versaleta...), ni caldran símbols per a marcar encavalcaments o una disposició del text determinada dins la pàgina. El fitxer resultant de la transcripció, sincronització i segmentació amb ELAN tindrà un format estàndard XML TEI (Text Encoding Initiative - TEI Consortium, 2017) (vg. Esplà et al. 2018).

Les dues transcripcions seran complementàries. La transcripció fonooortogràfica és a mig camí entre la *transcripció discursiva* que incorpora molts elements del context (Du Bois et al. 1991, Bladas 2009: §4) i la *transliteració ortogràfica*. Aquesta capa de transcripció permet tenir una versió molt pròxima a la fonètica del text oral i a les característiques morfosintàctiques, però no entrarà en detall a marcar aspectes prosòdics com l'entonació. És a dir, es tracta d'una transcripció intermèdia-estreta. La segona capa de transcripció, l'ortogràfica estandarditzada, és una versió d'acord amb l'ortografia convencional i la llengua normativa que facilitarà les tasques de segmentació, lematització i anotació. En aquest cas, doncs, es tracta d'una transcripció ampla o una *transliteració*. Cada parlant tindrà, doncs, aquestes dues capes de transcripció, dues versions (més o menys estretes) del mateix text. Amb tot, totes dues transcripcions sí que incorporaran alguns aspectes discursius que es puguen marcar amb la puntuació, s'indicaran les pauses, les interjeccions o altres sons (vg. §7).

En aquest protocol de transcripció presentem la segmentació de la llengua oral (§2), els criteris de la transcripció fonooortogràfica (§3), els criteris de la transcripció ortogràfica estandarditzada (§4), els criteris de puntuació (§5), els criteris tipogràfics (§6), els aspectes prosòdics (§7) i el tractament de la privacitat (§8).

2. Segmentació de la llengua oral

La segmentació del discurs en unitats de transcripció l'elaborarem semiautomàticament. Primer, amb el programa ELAN (vg. Esplà et al. 2018) obtindrem una segmentació automàtica en què s'aprofiten les pauses per a marcar una unitat o segment. Aquesta segmentació serà reajustada pel transcriptor per assegurar-nos que el tall es produeix coincidint en una pausa. El transcriptor també ajustarà manualment cada segment a la línia del parlant que té la paraula, és a dir, cada segment estarà situat a la línia del parlant en qüestió. Evidentment, hi haurà segments solapats sempre que hi haja més d'un parlant intervenint al mateix moment.

Cada segment serà més o menys equivalent a la unitat tonal o unitat prosòdica (Prieto & Cabré 2013). La segmentació de cada unitat tonal no serà del tot exhaustiva perquè és molt difícil determinar les característiques tonals amb una o dues audicions (Edwards 1993 *apud* Bladas 2009) (vg. §7). Tanmateix, el transcriptor intentarà acostar-s'hi i aprofitar el silenci de cada parlant per a acabar un segment.

3. Criteris lingüístics de la transcripció fonooortogràfica

La transcripció fonooortogràfica pretén mostrar la realització fonètica dialectal del text oral utilitzant les convencions de l'ortografia catalana. Com a criteri general, ens acostarem al màxim a la fonètica real, però posarem com a límit un criteri d'intel·ligibilitat, és a dir, que la distància amb l'ortografia estàndard no en dificulti la lectura, ja que tampoc pretenem fer una

transcripció fonètica rigorosa ni usarem l'Alfabet Fonètic Internacional (AFI). Així per exemple, marquem la caiguda de la *d* intervocàlica (*vesprà*), però no l'elevació lingual o obertura de totes les vocals (no transcriurem *còsa*, sinó *cosa*, tot i que la *o* siga oberta), ni el ieisme o el betacisme.

Fonètica

La transcripció fonoortogràfica **representa** els fenòmens fonètics següents, quan apareixen:

- L'harmonia vocàlica: *terre* 'terra'; *porto* 'porta'.
- L'elevació de les vocals *e* i *o* quan porten accent gràfic. És a dir, distingirem gràficament la realització *telèfon* de *teléfon*, però no *pou* ['pow] de *pou* ['pɔw] o *biblioteca* [biblio'teca] de *biblioteca* [biblio'teca] perquè no porten accent gràfic.
- Les neutralitzacions de vocals: *asperar* 'esperar', *calandari* 'calendari', *llançol* 'llençol'.
- Les diverses realitzacions vocàliques de l'auxiliar *haver*: *Jo ha fet* 'jo he fet'.
- La preposició *a* l'escriurem *a*, *ad*, *an* segons com la pronuncie el parlant: *Li ho donaré ad ell*.
- La caiguda de la *d*, *g* o *v* intervocàliques: *vesprà* 'vesprada', *juar* 'jugar' o *adaptàem* 'adaptàvem'.
- L'ensordiment de la [z] del mot *zero*: *sero* o la sonorització de *zenzillo* 'senzill'. En canvi, els ensordiments sistemàtics de les sibilants no els marcarem.
- Altres caigudes de vocals o consonants més locals (*nessitar* 'necessitar', *metla* 'ametla', *tallaiïns* 'tallarins', *carreó* 'carreró'..)
- L'emudiment de la *-r* final: *di* 'dir', *comprà* 'comprar', *corredó/correó* 'corredor'. I, per tant, accentuarem d'acord amb les regles ortogràfiques vigents: *comprà*.
- Fenòmens fonètics diversos: *ottubre* 'octubre', *quidrar/quirdar* 'cridar', *bragó* 'braó', *a von* 'a on', *moixca* [moxca o moixa] 'mosca'. També transcriurem els casos de variants fonètiques incorporades en diccionaris normatius: *flare* 'frare',
- La caiguda de l'auxiliar de perfet *haver*: *li dit* 'li he dit', *li's fet* 'li has fet', *li'm dit* 'li hem dit'.
- Les diverses realitzacions fonètiques del pronom *ho*, [o], [w], [ew] o [aw], per exemple: *voler-ho*, *valora-hu*, *hu sé*, *hau sé*, *heu sé*, *t'heu dic*, *heu agarre*, *hau agarre*...
- La contracció de la preposició *per* (*a*) que no coincidisca amb l'ortografia: *pa tu* ('per a tu'), *pa fer-ho* ('per a fer-ho'). Davant els adverbis demostratius usarem l'apostrofació: *p'ahí* o *pe'hí* (apòstrof sense espai ja que té un únic accent).
- La palatització d'*haver-hi*: *no ny'ha* 'no n'hi ha' (apòstrof sense espai ja que té un únic accent).
- Indicarem la reducció de la vocal de suport de l'article definit [l]: *ara 'ls ulls* 'ara els ulls', *arribarà 'l dia* 'arribarà el dia'. També la reducció de la [l] en el plural dels articles: *es troncs* 'els troncs'.
- Les elisions les marcarem amb un apòstrof: *onz' anys*, *un' hora*, *la mateix' hora*, *sens' ou*; *eix' home* ja que això ens permetrà distingir-ho de la pronúncia sense l'elisió: *onze anys*, *una hora*... (deixarem un espai per separar les dues paraules).

Hi ha una sèrie de trets propis de certs dialectes (o idiolectes) que **no representem** amb una grafia diferent de l'ortografia estàndard:

- L'obertura de les vocals *e* i *o* quan no porten accent gràfic.

- L'ensordiment de les sibilants sonores: *casa, dotze, metge, pluja*. Tampoc marcarem els canvis produïts per la fonètica sintàctica: *peix i carn, cases altes*.
- La sonorització en fonosintaxi de les oclusives sordes: *cin[g] o sis; se[d] o huit*.
- La presència/absència de la iod davant de la palatal fricativa sorda i sonor (*caixa/caxa, bajoca/baijoca, corretja/correija, pujar/puijar, metge/meige*).
- Les aspiracions de la *s*: *és que*.
- El ieisme: *llavi, palla, pell*
- El betacisme: *vinc, vols*.
- La distinció entre fricatives [ʒ] i africades [dʒ]: *Jaume*.
- Les diferents realitzacions fonètiques del pronom *jo* [io], [dʒo]... Sempre transcriurem *jo*.
- La realització de les oclusives fricatives: *ceba, pagar...*
- La tonicitat o atonicitat de l'auxiliar de perfet: *ha fet* ['a 'fet] o [a 'fet].
- El reforç de les consonants finals: [pont^ɾ] ho transcriurem *pont*.
- La caiguda de les oclusives finals: *pont, alt, camp, alts, malalts*.
- La palatalització del grup *-ts* final: *acabats, tots*.
- No marcarem les sinalefes perquè és un fenomen molt general: *una amiga*.

Morfosintaxi

En general, transcrivim tots els trets morfosintàctics del discurs del parlant. Alguns exemples de morfologia nominal i verbal són:

- L'article *lo* i l'article neutre *lo*: *lo llibre, lo que dius, lo bonic que és*.
- L'article salat: *sa llibreta*.
- Els demostratius *este, eixe* i *aquell* es transcriuen mantenint la forma no reforçada i reflectint la realització fonètica: *ixe xic, esta dona, est' home*.
- Les realitzacions morfològiques dialectals dels pronoms forts com *voatros* 'vosaltres', i dels pronoms febles com *mos, nos*, etc.: *me fa* 'em fa', *mos diu* 'ens diu', *se n'anem* o *mo n'anem* 'ens n'anem' *comprar-mo'n* 'comprar-nos-en', *compra-lo* 'compra'l'.
- La preposició *amb* la transcrivim *en* o *an*, d'acord amb la pronúncia: *an ell, en Maria*. No marcarem si la *n* s'assimila a la bilabial de la paraula següent: transcriurem *an pa* 'amb pa' i no, *am pa*.
- La flexió verbal mostrarà la forma usada: *perc* 'perd', *vega* 'veja', *ell cantave* 'cantava', *ell porte* 'porta', *servixca* 'servisca', *crexca* 'cresca', *vixcut* 'viscut'.
- Independentment que les formes verbals siguin normatives o no, reflectirem la forma usada pel parlant. Per exemple, en alguns verbs de la II conjugació, els parlars valencians no tenen la [j] antihiàtica: *féem, caem, veeu...*
- El sufix *-ea*: *vellea*.
- Duplicació del clític de datiu: *li vaig dir a ma mare*
- Altres pleonasmes: *n'hi havia dos cases*.
- La preposició *a* davant complement directe: *He vist a ma mare al mercat*.

Lèxic

Independentment de l'origen del mot o de les recomanacions prescriptives escriurem tots els mots sense modificar-los: *mensatge* 'missatge'. Si són mots dialectals utilitzarem com a referència el DCVB: *setiet* 'estalvis', *xicotiu* 'molt petit'. Si no apareixen en el DCVB (p. ex., *xicorrotiniu*) els transcriurem seguint els criteris fonètics i morfològics descrits.

Els manlleus els transcriurem usant l'ortografia catalana, independentment de si ha estat normalitzat pel TERMCAT, el Porterval, els diccionaris normatius (DIEC, DNV) o el GDLC: *tetxo* (esp. *techo*), *unya* (esp. *uña*), *sumo* (esp. *zumo*), *calsonsillos* (esp. *calsonillos*),

uàssap, uassap, uasap, uatzap... (ang. *whatsapp*), *quesso* (esp. *queso*), *trage* (esp. *traje*), *antonces* (esp. *entonces*), *limpio* (esp. *limpio*), *tuit* (ang. *twit*). Les paraules més consolidades en la llengua col·loquial presenten els trets de la flexió i la derivació catalanes (*unyes, sumet, tetxar, quesset*). Aquests casos, doncs, els que transcriurem amb l'ortografia catalana i els considerem manlleus adaptats.

En el cas de manlleus que tinguen sons que són aliens a la fonètica catalana, és a dir que no han estat adaptats, utilitzarem la grafia de l'idioma original (castellà, anglès, francès, etc.), tot i que no coincidisca amb la fonètica catalana: *jauja, zumo, jefe, jamón* o *jamó, heavy, traje, prêt-à-porter, voilà...* Aquests exemples són manlleus no adaptats.

En el cas dels manlleus no adaptats haurem d'afegir una anotació, és a dir, escriurem una etiqueta abans de la paraula afectada (<ManlleuNoAdaptat>) i una després (</ManlleuNoAdaptat>), sense deixar-hi espais en blanc. Vegeu l'exemple següent:

```
<ManlleuNoAdaptat>botellón</ManlleuNoAdaptat>
<ManlleuNoAdaptat>piscina</ManlleuNoAdaptat> [pronúncia castellana]
<ManlleuNoAdaptat>iPad</ManlleuNoAdaptat>
<ManlleuNoAdaptat>gauche divine</ManlleuNoAdaptat>
```

En el cas de les etiquetes és molt important que estiguen escrites correctament (fixeu-vos-hi que no hi ha espais, que hi ha majúscules, que no hi ha accents, etc.).

Les xifres s'escriuran seguint l'ortografia convencional i respectant la fonètica del parlant: *quatre-cents sixty-six-huit*. Segueix aquest criteri, la transcripció del nom de les hores (*són les onze i quart*) i el nom de les lletres, siguen manlleus o no (*ve, uve, jota, hatxe, be*)...

4. Criteris lingüístics de la transcripció ortogràfica estandarditzada

La transcripció ortogràfica estandarditzada s'elabora a partir de la transcripció fonooortogràfica. Com que l'objectiu és tenir una transcripció apta per a una tokenització i anotació morfològica ràpida i efectiva, en aquesta capa transcriurem usant la norma ortogràfica i morfològica en la majoria dels casos.

Fonètica

No representarem els fenòmens fonètics en aquesta transcripció. A diferència de l'anterior, en aquest cas tots **els fenòmens següents no els transcriurem** i usarem la forma ortogràfica normativa (OIEC i la GNV):

- L'harmonia vocàlica: *terra* o *porta*.
- L'elevació de les vocals *e* i *o* quan porten accent gràfic. Seguirem l'ortografia: transcriurem *telèfon*, independentment de la pronúncia [te'lefon] o [te'lefon].
- En català occidental hi ha alguns mots, que provenen del llatí amb una *Ē*, que han evolucionat a una [e] mentre que en català oriental tenen una [ɛ] o [ə]. Quan aquests mots tenen aquesta vocal en posició que ha de ser accentuada, d'acord amb les normes generals, optarem per l'accentuació general, amb una *e* oberta (*francès*) i no la que reflecteix la pronúncia occidental (*francés*). Els contextos més habituals són: terminacions *-es* de gentilicis (*anglès*), participis (*admès, compromès*) i adjectius (*cortès*); numerals ordinals acabats en *e* (cinquè) i alguns substantius (*café, comitè*); la terminació *-en* de la tercera persona del plural del present d'indicatiu d'alguns verbs de la II conjugació (*aprèn, comprèn, depèn*), llevat de casos en què aquesta

terminació està precedida de la consonant *t* o *c* (*atén, entén, pretén, encén*); els infinitius acabats en *-eixer* (*conèixer*) i *-encer* (*convèncer*); la segona i tercera persona del plural dels imperfects d'indicatiu amb accent al radical (*fèiem, fèieu*).

- Les neutralitzacions de vocals: *esperar, calendari, llençol*.
- Les diverses realitzacions vocàliques de l'auxiliar *haver*: *jo he fet*.
- La preposició *a* l'escriurem sempre *a* i no *ad, an*, independentment de la pronúncia del parlant: *Li ho donaré a ell*.
- La caiguda de la *d, g* o *v* intervocàliques: *vesprà* escriurem *vesprada*, *juar* escriurem *jugar* o *adaptàem* escriurem *adaptàvem*.
- L'ensordiment de la [z] del mot *zero* no la transcriurem així *sero*, sinó amb l'ortografia corrent, *zero*. El mateix per a altres mots com *zenzillo* 'senzill', el transcriurem *senzillo*.
- Altres caigudes de vocals o consonants més locals: *nessitar* escriurem *necessitar*, *metla* escriurem *ametla*, *tallaiñs* escriurem *tallarins*, *carreó* escriurem *carreró*, etc.
- L'emudiment de la *-r* final: *dir, comprar, corredor*.
- Fenòmens fonètics diversos (com *ottubre, flare, quidrar/quirdrar, bragó, a von*, etc), els transcriurem segons la norma (*octubre, frare, cridar, braó, a on*).
- La caiguda de l'auxiliar de perfet (*li dit, li's fet, li'm dit*): *li he dit, li has fet, li hem dit*.
- Les diverses realitzacions fonètiques del pronom *ho* [o], [w], [ew] o [aw] (com ara *hau sé, heu sé, t'heu dic, heu agarre, hau agarre, [w] agarre, compra-[w], comprar-[o]*): *ho sé, t'ho dic, ho agarre, compra-ho, comprar-ho*.
- Les contraccions de preposicions que no coincidisquen amb l'ortografia, com *pa tu, pe mi, p'ahí...* les reconstruirem d'acord amb la preposició estàndard: *per a tu, per a mi, per ahí*.
- La palatització d'*haver-hi, no ny'ha*, la transcriurem amb l'ortografia corrent: *no n'hi ha*.
- No indicarem la reducció de la vocal de suport de l'article definit [l] i per tant escriurem *ara els ulls* o *arribarà el dia*. També escriurem la *l* de l'article definit en plural encara que en la pronúncia caiga (*es troncs*): *els troncs*.
- No indicarem les elisions: *onze anys, una hora, la mateixa hora, sense ou, eixe home...*

Tal com ja hem dit en la transcripció fonoortogràfica, en aquest cas tampoc no representarem aquests fenòmens fonètics:

- L'obertura de les vocals *e* i *o* quan no porten accent gràfic.
- L'ensordiment de les sibilants sonores: *casa, dotze, metge, pluja*. Tampoc marcarem l'ensordiment del mot *sero* 'zero'. Tampoc marcarem els canvis produïts per la fonètica sintàctica: *peix i carn, cases altes*.
- La sonorització en fonosintaxi de les oclusives sordes: *cin[g] o sis; se[d] o huit*.
- La presència/absència de la iod davant de la palatal fricativa sorda i sonor (*caixa/caxa, bajoca/baijoca, corretja/correija, pujar/puijar, metge/meige*)
- Les aspiracions de la *s*: *es que*.
- El ieisme: *llavi, pell*.
- El betacisme: *vinc, vols*.
- La distinció entre fricatives [ʒ] i africades [dʒ]: *Jaume*.
- Les diferents realitzacions fonètiques del pronom *jo* [io], [dʒo]... Sempre transcriurem *jo*.
- La realització de les oclusives fricatives: *ceba, pagar...*
- La tonicitat o atonicitat de l'auxiliar de perfet: *ha fet* ['a 'fet] o [a 'fet].

- El reforç de les consonants finals: [pont^ɾ] ho transcriurem *pont*.
- La caiguda de les oclusives finals: *pont*, *alt*, *camp*, *alts*, *malalts*.
- La palatalització del grup *-ts* final: *acabats*, *tots*.
- Les sinalefes: *una amiga*.

Morfosintaxi

En general, no transcriurem tots els trets morfosintàctics del discurs del parlant, sinó que usarem la varietat estàndard. Alguns exemples de morfologia nominal o verbal són:

- Els demostratius *este*, *eixe* i *aquell* es transcriuen mantenint la forma no reforçada, però no mantindrem la realització fonètica. Així no transcriurem *ixe xic* o *est' home*, sinó *eixe xic* o *este home*.
- Les realitzacions morfològiques dialectals dels pronoms forts com *voatros* i dels pronoms febles com *mos*, *nos*, etc. les escriurem d'acord amb la norma: *vosaltres*, *ens*, *us/vos*. En lloc de *se n'anem* o *mo n'anem* escriurem *ens n'anem*, en lloc de *comprar-mo'n* escriurem *comprar-nos-en* i en lloc de *compra-lo* escriurem *compra'l*.
- Les diferents realitzacions de la preposició *amb* (*en* o *an*) no les marcarem: *amb ell*, *amb Maria*.
- La flexió verbal estàndard. En lloc de *perc* escriurem *perd*, en lloc de *vullc* escriurem *vull*, en lloc de *cusc* escriurem *cus*, en lloc de *vega* escriurem *veja*, en lloc de *ell cantave* escriurem *ell cantava*, en lloc de *servixca* escriurem *servisca*, etc.
- En alguns verbs de la II conjugació, els parlars valencians no tenen la [j] antihiàtica: *fèem*, *caem*, *veeu* en lloc de *fèiem*, *caiem*, *veieu*. Tot i que pot considerar-se la doble flexió en alguns casos també pròpia de l'àmbit formal, mantenim el model més convergent. Per tant, aquests casos els transcriurem per la forma amb la [j]: *fèiem*, *caiem*, *veieu*...

En canvi, altres trets dialectals sí que els conservarem perquè són fàcils d'anotar:

- L'article *lo* i l'article neutre *lo*: *lo llibre*, *lo bonic*, *lo d'ell*, *lo que passa*.
- Article salat: *sa llibreta*.
- Duplicació del clític de datiu: *li vaig dir a ma mare*
- Altres pleonasmes: *n'hi havia dos cases*.
- La preposició *a* davant complement directe: *He vist a ma mare al mercat*.
- El sufix *-ea*: *vellea*.

Lèxic

Bàsicament, seguim els criteris de la transcripció fonooortogràfica, amb algunes excepcions. En general, els mots dialectals els transcriurem d'acord amb el DCVB (o el DNV): *sitiet* o *setiet* 'estalvis', *xicotiu* 'molt petit'. En algun cas, però, els adaptarem a l'estàndard: *mensatge* 'missatge'.

A diferència del criteri de la transcripció fonooortogràfica, els manlleus que han estat adaptats pel TERMCAT o Porterval, els adaptarem usant la forma proposada: *whatsapp*... En canvi, els mots no normalitzats, seguint el criteri de la transcripció fonooortogràfica, els transcriurem usant l'ortografia catalana: *tetxo* (esp. *techo*), *unya* (esp. *uña*), *sumo* (esp. *zumo*), *calsonillos* (esp. *calsonillos*), *quesso* (esp. *queso*), *trage* (esp. *traje*), *antonces* (esp. *entonces*), *límpio* (esp. *limpio*), *tuit* (ang. *twit*). Les paraules més consolidades en la llengua col·loquial presenten els trets de la flexió i la derivació catalanes (*unyes*, *sumet*, *tetxar*, *quesset*). Aquests casos, doncs, els que transcriurem amb l'ortografia catalana.

En el cas de manlleus que tinguen sons que són aliens a la fonètica catalana, és a dir que no han estat adaptats, utilitzarem la grafia de l'idioma original (castellà, anglès, francès,

etc.), tot i que no coincidisca amb la fonètica catalana: *jauja, zumo, jefe, jamón o jamó, heavy, traje, prêt-à-porter, voilà...* Aquests exemples són manlleus no adaptats.

Tal com hem vist a §3, marcarem els manlleus no adaptats amb les etiquetes <ManlleuNoAdaptat> i </ManlleuNoAdaptat>.

Les xifres s'escriuran seguint l'ortografia estàndard, independentment de la fonètica del parlant: *quatre-cents seixanta-vuit*. Les hores (*les onze i quart*) i el nom de les lletres seguiran el mateix criteri: *ve, jota, hac, be* (i no *be alta*).

5. Puntuació

La puntuació és un sistema gràfic molt ric per a la llengua escrita que vol organitzar el discurs, separar les oracions o constituents oracionals i facilitar-ne la lectura. Per tant, la puntuació no serveix per a marcar les pauses, sinó que fa altres funcions. Usar la puntuació per a la transcripció de la llengua oral és, doncs, aplicar-la a un àmbit que li és aliè, en principi. Tanmateix, en alguns casos farem una reinterpretació de la puntuació i l'usarem.

Els criteris són parcialment diferents en la transcripció fonootogràfica i en la transcripció ortogràfica. Per tant, vegem-los separadament:

Transcripció fonootogràfica

En principi, les pauses s'indiquen amb la segmentació (vg. §2). Ara bé, pot passar que dins d'un segment hi haja una pausa. En aquest cas, usarem els símbols:

/ pausa breu

// pausa llarga

Per tant, en la transcripció fonootogràfica, no marcarem les pauses amb els símbols que utilitza la fonètica –com ara (|) o (||)– ni amb els signes de puntuació, com el punt (.), la coma (,), els dos punts (:), els guions (–), el punt i coma (;) o els parèntesis.

Transcripció ortogràfica estandarditzada.

En la transcripció estandarditzada sí que indicarem les pauses amb punts i comes, com expliquem tot seguit. Ara bé, tampoc no utilitzarem els altres signes de puntuació: els dos punts (:), els guions (–), el punt i coma (;) o els parèntesis.

El punt (.)

El punt serveix per a marcar la **pausa llarga** i/o el **canvi de tema**. Els usos més habituals són els següents, però només usarem el punt en aquests casos (o d'altres) sempre que hi haja pausa o canvi de tema:

- Al final d'una oració que reflexiona sobre una mateixa idea o quan hi ha un canvi de tema. *Em va dir que estava molt cansat i que no vindria. Me'n vaig a comprar.* Cal tenir en compte que el concepte d'oració en la llengua oral no és tan clar com en l'escripta i, per això, hi poden haver contextos que facen de mal dir. Recorda també que no hi ha paràgrafs i, per tant, tampoc hi ha punts i a part.
- Al final del text, per a indicar-ne l'acabament.
- Per indicar que hi ha dues oracions diferents, però que tenen algun vincle en relació amb el contingut, ortogràficament s'usa el punt i coma. En la transcripció usarem el punt: *Al final han tornat a les tres de la matinada, de veres. Massa tard, no trobes?*

- Per introduir una oració que explica millor l'anterior habitualment s'usen els dos punts. En la transcripció no podem usar els dos punts, per tant, en aquest cas usarem el punt (.) o la coma (,), segons la llargada de la pausa: *Al final havia de caure malalt. Tot no es pot fer alhora.*

La coma (,)

En general, usem la coma per a marcar la **pausa breu**. Els casos més habituals són els següents, però sols usarem la coma en aquests casos (o d'altres) si el parlant fa una pausa breu:

- Per separar els components d'una enumeració: *he comprat pomes, creïlles, i cebes*. Si hi ha pausa breu entre *creïlles* i la conjunció *i*, també hi posarem coma. Recordem que no és un ús ortogràfic.
- Per iniciar una enumeració, si hi ha pausa breu: *he comprat de tot, pomes, creïlles, i cebes*.
- Per introduir un incís o aclariment, és a dir, un comentari secundari que es podria llevar sense que el discurs se'n ressentirà: *quan vam arribar-hi, és que era molt tard, estava tancat*. No usarem, doncs, guions o parèntesis.
- Per marcar un canvi de l'ordre lògic de l'oració: *La blanca, m'he comprat al final*.
- Per indicar que algun element s'ha omès perquè es pot recuperar a partir del discurs anterior: *Hui de matí ha perdut la cartera i a la vesprada, les claus*.
- Després d'una oració subordinada adverbial, si se situa davant de l'oració principal: *Quan he vingut al matí, no hi havia cotxes aparcats. Si tens temps, repassa les notes que has pres*.
- Sovint, després d'alguns connectors com *en canvi, per un costat, en definitiva*, etc.
- Entre el subjecte i el predicat: *La faena que m'has encomanat, és molt difícil*. En la llengua escrita, aquest ús de la coma es considera incorrecte. En la transcripció, però, l'usarem si el parlant fa una pausa.
- Per separar elements dins d'una oració, encara que ja hi haja elements separats per comes: *El president va assenyalar, en primer lloc que havien de llançar la brossa abans de les onze de la nit, que havien de buidar les bústies de propaganda cada dos dies, però no llançar els papers en el portal, que no fumaren en l'ascensor i, finalment, que pagaren la contribució*.
- Per marcar que comença la reproducció de les paraules exactes d'algú altre, una citació: *Va agafar el micròfon i va dir, He decidit que aquesta nit esteu tots convidats*.

Altres signes de puntuació per a les dues transcripcions

A més a més de marcar les pauses, en totes dues transcripcions usarem alguns signes de puntuació per a indicar alguns aspectes prosòdics: la interrogació (?), l'exclamació (!) i els punts suspensius (...). Vegem-ne els usos:

Interrogació (?)

Usarem el signe d'interrogació al final d'una oració per indicar que fem una pregunta directa: *Què passa? Anem a comprar?*

Exclamació (!)

Usarem el signe d'exclamació per expressar gràficament sorpresa o èmfasi: *Au! Ves-te'n a pastar fang!*

Punts suspensius (...)

Usarem els punts suspensius en el mateix sentit que la llengua escrita: per marcar que una **oració és inacabada**, que hi ha una interrupció, que l'oració es deixa en suspens o que una enumeració no s'ha acabat: *És que ell és molt...*

Cometes (“”)

No usarem les cometes. L'estil directe o les citacions no les indicarem amb la puntuació, sinó amb un etiquetatge específic (vg. §7).

6. Criteris tipogràfics

Com ja hem dit, no utilitzarem la negreta, la cursiva, les versaletes, el subratllat ni cap altre recurs tipogràfic.

Majúscules i minúscules

A diferència de la llengua escrita, restringirem l'ús de la majúscula a casos molt clars i molt concrets:

- La primera lletra de l'inici del discurs de cada parlant.
- En el cas de la transcripció fonoortogràfica, usarem la majúscula després d'una pausa llarga amb canvi de tema.
- En el cas de la transcripció ortogràfica, usarem la majúscula després de punt.
- Per a iniciar un discurs reportat, encara que abans hi haja una coma (que indica pausa breu): *Va agafar el micròfon i va dir, He decidit que aquesta nit esteu tots convidats.*
- Els noms propis: antropònims (*Enric Valor*), topònims (*el Montgó*), els noms d'institucions, organismes o empreses (*Generalitat, la Caixa*). També quan siguen acrònims: *Renfe, Unicef*.
- Entitats religioses: *Déu, Esperit Sant*.

Així doncs, escriurem en minúscula els altres casos i, sobretot, en cas de dubte, prioritzarem l'ús de la minúscula: *ajuntament, diputació, estat...*

7. Aspectes prosòdics i altres

Entonació

El corpus *Parlars* vol ser un corpus per a estudis de diversos tipus, especialment gramatical. Per això, l'estudi de la prosòdia no és l'objectiu principal. També volem que siga un corpus extens i robust. Així doncs, el detall en la transcripció prosòdica ha de reduir-se al mínim. Tot i això, tal com hem vist en §5, marcarem mínimament els aspectes prosòdics, ja que usarem les exclamacions, interrogacions ortogràfiques i els punts suspensius per a marcar l'entonació descendent (*No tens raó*), l'ascendent (*Vindràs demà?*) i el manteniment (*El que em vas explicar ahir...*), respectivament.

Allargament de vocals

En la transcripció fonoortogràfica, marquem amb la repetició de les vocals pertinents els allargaments de les vocals finals que sovint es donen en el discurs oral per a manifestar sorpresa i exclamació. L'allargament podrà ser de dues, tres o quatre vegades el so allargat, segons la duració:

- *Quèè?, Quèèè? o Quèèèè?*
- *Síiii.*
- *Volss... vindre?*

En la transcripció ortogràfica no indicarem l'allargament:

- *Què.*
- *Sí.*
- *Vols... vindre?*

Pauses i encavalcaments

Ja hem vist com es marquen les pauses (vg. §3 i §5). Si les pauses són molt llargues o molt breus es podrà observar en l'enregistrament i no amb la transcripció. No es marcaran tampoc els encavalcaments conversacionals perquè cada parlant té una capa d'anotació individual i, per tant, hi poden haver superposaments en el mateix moment temporal (vg. Esplà et al. 2018).

Interjeccions i sons paralingüístics

En el registre col·loquial solem trobar diversos recursos lingüístics discursius i expressius que doten d'eficàcia el missatge i que podem englobar dins de l'etiqueta general de les interjeccions (Cuenca 2002). La transcripció d'aquests recursos és, sovint, complexa. Per això, seguirem l'ortografia adaptada que proposa Cuenca (2002) i la completarem amb la proposta de Riera-Eures & Sanjaume (2002 i 2010) i el llistat de 163 interjeccions que podem recuperar en la cerca avançada del DNV. En cas de ser necessari, emprarem la grafia convencional *h* per a marcar aspiracions tant en els casos més estandarditzats, com *ha-ha-ha* o *ehem* com els que no ho són tant, com *ahà* o *ha*.

Les interjeccions més generals són:

Valor	Forma
alegria, enuig	<i>xe, xi</i>
apel·lació	<i>txit, psit</i>
assentiment, corroboració o comunicació que se segueix la conversa	<i>mhm, hu-hu</i>
avís o prevenció	<i>ui-ui-ui</i>
criada, avís o salutació	<i>ei, ep, ie</i>
demanda de conformitat	<i>eh, tat, vitat</i>
desacord	<i>ps</i>
dolor	<i>ai, au, oi, ui</i>
èmfasi	<i>fú, buf</i>
estranyesa	<i>ai</i>
fàstic	<i>ec, ecs, uà, uf, uix</i>
fruïció, desig	<i>hum</i>
frustració o desacord respecte al que s'ha dit	<i>ntx</i>
indiferència o menyspreu	<i>pse</i>
inici o canvi de tema de conversa i altres usos	<i>ah</i>

menyspreu	<i>tx</i>
objecció o rectificació	<i>ep, uep</i>
omplidor de la parla que permet conservar el torn mentre es pensa què dirà a continuació	<i>e..., mm</i>
opressió, aclaparament	<i>uf</i>
per a saludar, cridar l'atenció o demanar silenci	<i>ts</i>
sorpresa, desil·lusió o contrarietat	<i>uei</i>
sorpresa, incredulitat	<i>ai, ca, ha, hu</i>

Taula 1. *Interjeccions* (basada en Bldas 2009, DNV, Riera-Eures & Sanjaume 2002 i 2010)

D'altra banda, transcriurem les altres interjeccions pròpies o impròpies (normalment localismes), siguin catalanes o manlleus, que no apareixen en els reculls esmentats adés seguint els criteris descrits (vg. §3 i §4): *arrea, ira, baia, equiliquà, iò, voilà*, etc.

En la capa de lèxic, anotarem el sentit d'aquestes formes, si no coincideixen amb les previstes i el transcriptor considera que cal aclarir-ho.

Les interjeccions les marquem amb una marca morfològica en la capa de la categoria gramatical.

Onomatopeies

Per a la transcripció de les onomatopeies seguirem el llistat de 92 onomatopeies present en el DNV i el completarem, en cas de necessitat, amb el que conté Riera-Eures & Sanjaume (2002 i 2010). En cas que aparega en el discurs algun element onomatopèic no present en els repertoris assenyalats, el transcriurem seguint els criteris descrits adés (vg. §3 i §4).

Les onomatopeies han d'anar marcades amb una etiqueta específica abans i després de la paraula, com en l'exemple següent (sense deixar-hi espais):

<Onomatopeia>bub-bub</Onomatopeia>

Riures i altres sons

Els sons provocats pels parlants (riures, esternuts, tos, respiració, sospirs, xiulet...) o sons externs (portes, cotxes...) no els marquem en les transcripcions, tot i que se sentiran en l'àudio.

Paraules o fragments incomprensibles

Les paraules o fragments incomprensibles els marcarem així: <Incomprensible/>. Aquest marcatge no té una etiqueta d'inici i una de final, sinó que és una etiqueta única que indica que hi ha una paraula o fragment que no podem desxifrar.

Si la paraula o fragment és difícil d'entendre però tenim una hipòtesi raonable ho marcarem així:

<Dubte>paraula o paraules dubtoses</Dubte>

Estil directe

Tant en la transcripció fonootogràfica com en l'estandarditzada, els fragments en estil directe o les citacions els marcarem així amb les següents etiquetes: <EDirecte> i </EDirecte>. Exemple:

Vam arribar a la plaça i fa <EDirecte>És que mai ve ningú. Quin desastre!</EDirecte>

Topònims

Els topònims menors (partides, muntanyes locals, noms de carrers...) els marcarem amb les etiquetes següents:

<Lloc>Segària</Lloc>

Carrer <Lloc>Major</Lloc>

8. Participants i anonimització

L'entrevistador és un col·laborador del projecte que coneix els informants, els enregistra i intenta fer fluir la conversa. Quan es tracta d'una entrevista-monòleg, la seua intervenció tendeix a ser mínima. Tot i això, quan intervé hi haurà una capa amb la transcripció de les seues intervencions. L'abreviació serà E. Si n'hi ha més d'un, els indicarem així: E1, E2, E3...

Per motius ètics i de protecció de dades, els textos han d'anonimitzar els informants i altres referents humans que s'esmenten. De fet, l'enregistrament també amagarà aquests fragments. Per tant, quan s'esmente un informant, el substituïrem per una numeració. Cada parlant tindrà una numeració diferent als altres parlants que apareguen en el corpus i, per això, caldrà comprovar en cada ocasió quin ha estat el darrer número consignat. Per exemple, *Vicent Miralles Pons* serà 001, si és el primer informant del corpus. Si en aqueixa conversa, un parlant diu "Tu, Vicent, què trobes?", ho transcriurem per "Tu, <NomInterlocutor ID="001"/>, què trobes?".

També marcarem els malnoms o les referències identificables a altres persones amb unes etiquetes, la qual cosa ens permetrà anonimitzar el text. Vegem-ne uns exemples:

Com em va dir <Malnom>Quico el del Pla</Malnom>

Mira per ací ve <Malnom>Teresa la de la plaça del rellotge</Malnom>

Quan el parlant fa referència a persones de fora de la conversa també marcarem el nom amb una etiqueta:

M'ha dit <NomPropi>Batiste</NomExtern> que no vindrà

Referències bibliogràfiques

Albelda, Marta (2005): "Sistemas de transcripción de los corpus orales del español", Carrió Pastor, M. Luisa, *Perspectivas interdisciplinarias de la lingüística aplicada*, vol. 2, p. 381-387.

Alturo, Núria; Òscar Bladas, Marta Payà & Lluís Payrató (ed.) (2004): *Corpus oral de registres. Materials de treball*. Barcelona: Publicacions i Edicions de la Universitat de Barcelona.

Beltran, Vicent; Segura-Llopes, Carles (2017): *Els parlars valencians*. València: PUV.

- Bladas, Òscar (2009): *Manual de transcripció del discurs oral. Materials de treball*. Universitat de Barcelona.
- Boix-Fuster, Emili; Marina Àlamo Sala, Mireia Galindo Solé, Francesc Xavier Vila i Moreno (ed.) (2007): *Corpus de Varietats Socials. Materials de treball*. Barcelona: Publicacions i Edicions de la Universitat de Barcelona.
- Briz Gómez, Antonio & Grupo Val.Es.Co. (2002): “Corpus de conversaciones coloquiales”, *Anejo de la revista Oralía*, Madrid, Arco-Libros.
- Brugman, H.; Russel, A. (2004): *Annotating Multimedia/Multi-modal resources with ELAN*, Proceedings of the Fourth International Conference on Language Resources and Evaluation, p. 2065-2068, Lisboa, Portugal.
- CCCUB = *Corpus de Català Contemporani de la Universitat de Barcelona*. (<<http://www.ub.edu/ccub/>>)
- DCVB = A. M. Alcover & F. de B. Moll (1928-1962) *Diccionari Català-Valencià-Balear*, Palma, Editorial Moll, 10 vol. [<http://dcvb.iecat.net.>]
- DNV = Acadèmia Valenciana de la Llengua, *Diccionari Normatiu Valencià*. (<<http://www.avl.gva.es/lexicval/>>)
- DIEC = Institut d'Estudis Catalans, *Diccionari de la llengua catalana*, 2a edició.
- GDLC = *Gran Diccionari de la Llengua Catalana*, Enciclopèdia Catalana.
- Carrera-Sabaté, Josefina & Joaquim Viaplana (ed.): *Corpus Oral Dialectal (COD). Textos orals del nord-occidental*. Dipòsit Digital de la UB.
- COC = Payrató, Lluís & Núria Alturo (ed.) (2002): *Corpus oral de conversa col·loquial. Materials de treball*. Barcelona: Publicacions de la Universitat de Barcelona.
- Cuenca, Maria Josep (2002): “Els connectors textuais i les interjeccions” dins Solà, Joan (dir.) *Gramàtica del català contemporani*. Barcelona: Empúries, 3173-3237.
- De Benito Moreno, Carlota, Javier Pueyo & Inés Fernández-Ordóñez (2016): “Creating and designing a corpus of rural Spanish”, *Proceedings of the 13th Conference on Natural Language Processing (KONVENS 2016)*, p. 78-83. (<https://www.linguistics.rub.de/konvens16/pub/10_konvensproc.pdf>)
- Esplà, Miquel; Beltran, Vicent; Guardiola, M. Isabel; Montserrat, Sandra; Segura, Carles; Sentí, Andreu (Corvalc) (2018): *Tutorial d'ELAN per a la transcripció del corpus Parlars. Primera part*. València: MMedia, Universitat de València. (<<https://mmedia.uv.es/buildhtml/52147>>)
- Esplà, Miquel; Beltran, Vicent; Guardiola, M. Isabel; Montserrat, Sandra; Segura, Carles; Sentí, Andreu (Corvalc) (2019): *Tutorial d'ELAN per a la transcripció del corpus Parlars. Segona part*. València: MMedia, Universitat de València. (<<https://mmedia.uv.es/buildhtml/53640>>)
- Forcada, Mikel L.; Ginestí-Rosell, Mireia; Nordfalk, Jacob; O'Regan, Jim; Ortiz-Rojas, Sergio; Pérez-Ortiz, Juan Antonio; Sánchez-Martínez, Felipe; Ramírez-Sánchez, Gema; Tyers, Francis M. (2011): *Apertium: a free/open-source platform for rule-based machine translation*. Machine Translation (Special Issue on Free/Open-Source Machine Translation) 25:2, 127-144.
- GNV = Acadèmia Valenciana de la Llengua, *Gramàtica normativa valenciana*. (<<http://www.avl.gva.es/documents/31987/65233/GNV>>)
- Nissim, Malvina & Paola Pietrandrea (2017): “MODAL: A multilingual corpus annotated for modality”. In *Proceedings of the Fourth Italian Conference on Computational Linguistics (CLiC-it 2017)*.
- Nivre, Joakim (2008): “Treebanks”, dins Lüdeling, Anke & Kytö, Merja (ed.) *Corpus linguistics. An International Handbook*. Berlin / New York: Walter de Gruyter.

- Observatori de Neologia (2004): *Metodologia del treball en neologia: criteris, materials i processos*. Barcelona: Universitat Pompeu Fabra. Institut Universitari de Lingüística Aplicada.
- OIEC = Institut d'Estudis Catalans, *Ortografia catalana*. (<https://www.iec.cat/llengua/documents/ortografia_catalana_versio_digital.pdf>)
- Payrató, Lluís & Núria Alturo (ed.) (2002): *Corpus oral de conversa col·loquial. Materials de treball*. Barcelona: Publicacions de la Universitat de Barcelona.
- Payrató, Lluís (2010): *Pragmàtica, discurs i llengua oral*. Barcelona: UOC.
- Perea, Maria-Pilar & Joaquim Viaplana: *Corpus Oral Dialectal (COD). Selecció de textos*. Dipòsit Digital de la UB.
- Pons, Clàudia & Joaquim Viaplana (ed.) (2009): *Corpus oral dialectal (COD)*. Textos orals del balear. Dipòsit Digital de la UB.
- Porterval = Acadèmia Valenciana de la Llengua, *Portal Terminològic Valencià*. (<<https://www.avl.gva.es/lexicval/ptv/>>)
- Prieto, Pilar & Cabré, Teresa (coords.) (2007-2012): “Criteris bàsics de transcripció ortogràfica de l’Atles interactiu de l’entonació del català”, dins *Atles interactiu de l’entonació del català*. Pàgina web: <<http://prosodia.upf.edu/atlesentonacio/>>.
- Prieto, Pilar & Cabré, Teresa (coords.) (2013): *L’entonació dels dialectes catalans*. Barcelona: Publicacions de l’Abadia de Montserrat.
- Riera-Eures Manel & Margarida Sanjaume (2002): *Diccionari d’onomatopeies i mots de creació expressiva*. Barcelona: Edicions 62.
- Riera-Eures Manel & Margarida Sanjaume (2010): *Diccionari d’onomatopeies i altres interjeccions*. Vic: Eumo.
- Samper Padilla, A., C. Hernández & M. Troya (1998): *Macrocorpus de la norma lingüística culta de las principales ciudades del mundo hispánico (MC-NLCH)*. Las Palmas de Gran Canaria. Las Palmas: Servicio de Publicaciones de la Universidad de Las Palmas de Gran Canaria.
- Solà, Joan et al. (dir.) (2002): *Gramàtica del català contemporani*, Barcelona, Empúries.
- TEI Consortium, eds. *Guidelines for Electronic Text Encoding and Interchange*. [2017-11-29]. <<http://www.tei-c.org/P5/>>.
- TERMCAT = *Terminologia Catalana* (<<http://www.termcat.cat/ca/>>)
- Viaplana, Joaquim & Maria Pilar Perea (ed.) (2003): *Textos orals dialectals del català sincronitzats. Una selecció*. Barcelona: Promociones i Publicaciones Universitarias (PPU).
- Viaplana, Joaquim; Maria-Rosa Lloret, Maria-Pilar Perea & Esteve Clua (2007): *COD. Corpus Oral Dialectal*. Barcelona: Promociones i Publicaciones Universitarias (PPU).
- Veny, J. (2002 [1982]): *Els parlars catalans*. Palma: Editorial Moll.

Citació bibliogràfica d’aquest treball:

Beltran, Vicent; Esplà, Miquel; Guardiola, M. Isabel; Montserrat, Sandra; Segura, Carles; Sentí, Andreu (Corvalc) (2019): *Criteris per a la transcripció del corpus Parlars*. València: Universitat de València. Roderic.