

## El COC del CUB: un corpus per a l'estudi de la conversa col·loquial<sup>1</sup>

2 El Corpus de Català Contemporani de la  
Universitat de Barcelona (CUB)

### 1 Introducció

Dins el conjunt de la recerca pròpia de les ciències del llenguatge s'ha anat posant de manifest, en les dues últimes dècades, la necessitat de comptar amb corpus lingüístics, és a dir, compilacions organitzades de dades lingüístiques que facilitin la tasca dels investigadors en els diversos corrents d'estudi que tenen com a centre el llenguatge.

Aquesta necessitat s'ha fet palesa segurament d'una manera encara més evident en tots els camps relacionats amb l'anàlisi de la llengua oral, i en particular de la conversacional i de la informal, en què la recollida, l'emmagatzematge, el maneig i la consulta de les dades resulten, en termes generals, més complexos que en altres camps. La situació és perfectament aplicable al cas dels estudis sobre la llengua catalana, que no es poden beneficiar encara de la utilització corrent de corpus o arxius de dades orals.

L'objectiu d'aquest article és presentar un corpus que ha estat confeccionat amb la intenció d'oferir als estudiosos del llenguatge uns materials adequats per a l'anàlisi de la conversa col·loquial: el Corpus Oral de Conversa Col·loquial (COC) del Corpus de Català Contemporani de la Universitat de Barcelona (CUB). El COC és una part d'un projecte més ampli, el CUB, que aplega també altres corpus representatius de diverses varietats geogràfiques, socials i funcionals de la llengua catalana actual. Dediquem el segon apartat d'aquesta presentació a explicar l'estructura del CUB. El tercer apartat conté una exposició detallada de les característiques del COC.

El Corpus de Català Contemporani de la Universitat de Barcelona (CUB) és un arxiu de dades textuais i morfològiques representatives de diverses varietats de la llengua catalana actual que es troba en una fase avançada de constitució. L'arxiu està estructurat en set subcorpus independents i alhora interrelacionats: un de varietats geogràfiques (Corpus Oral Dialectal, COD), un de varietats socials (Corpus Oral de Varietats Socioculturals, COS) i cinc de varietats funcionals (Corpus Oral de Conversa Col·loquial, COC; Corpus Oral de Registres, COR; Corpus Oral de Publicitat, COP; Corpus d'Informatius Orals, CIO; Corpus Escrit de Català Actual, CECA).

La recollida i transcripció d'aquests materials és un dels objectius del projecte *La variació en el llenguatge. El català actual*, que es va iniciar l'any 1991 en el marc de la Secció de Lingüística del Departament de Filologia Catalana de la Universitat de Barcelona. Uns altres objectius són la preparació dels materials del corpus per tal de fer-los accessibles a la comunitat científica (digitalització, sincronització text-veu, confecció de bases de dades de materials i parlants, etc.) i la realització d'estudis específics sobre la variació en la llengua catalana, basats en les dades del CUB. Paral·lelament, els investigadors del projecte promouen la realització de congressos i seminaris sobre temes relacionats amb la variació lingüística (com les 7 edicions del *Col·loqui Lingüístic de la UB*, o les 9 jornades i seminaris de la *Xarxa temàtica La variació lingüística: dialectologia, sociolingüística i pragmàtica*). El grup, a més, forma part del *Centre de Referència en Enginyeria Lingüística* de la Generalitat de Catalunya, i participa en les activitats d'altres xarxes temàtiques (*Xarxa d'Estudis del Discurs*, *Xarxa de Lingüística Aplicada*), a més d'encapçalar el grup de recerca consolidat *Grup d'Estudi de la Variació* (GEV), que integra també altres equips de recerca dirigits per investigadors de la Secció de Lingüística Catalana de la UB.

Des dels seus inicis, han participat en aquest projecte un total de 22 professors universitaris, 8 becaris predoctorals (CIRIT i DGICYT) i 30 becaris del projecte i col·laboradors voluntaris de segon i tercer cicle. A més, un gran nombre d'estudiants, de segon i tercer cicle, han aportat al

<sup>1</sup> Aquest article s'ha elaborat en el marc del projecte *La variació en el llenguatge. El català actual*, del grup de recerca consolidat Grup d'Estudi de la Variació. La realització d'aquest projecte és possible gràcies al finançament de la DGICYT (PB97-0889) i de la CIRIT (SGR98-00041).

projecte materials que en molts casos han estat incorporats al CUB.<sup>2</sup> La infraestructura necessària, així com bona part de les tasques derivades dels objectius del projecte, han estat finançats per la UB, la CIRIT i la DGICYT.<sup>3</sup>

## 2.1 Estructura del CUB

El CUB va néixer a partir de l'interès per l'estudi de la variació en la llengua catalana i, particularment, en el català oral. Aquesta circumstància va condicionar, des dels inicis, la configuració del corpus, que és, predominantment, un corpus de llengua oral organitzat amb una estructura modular que es presenta en la figura 1.

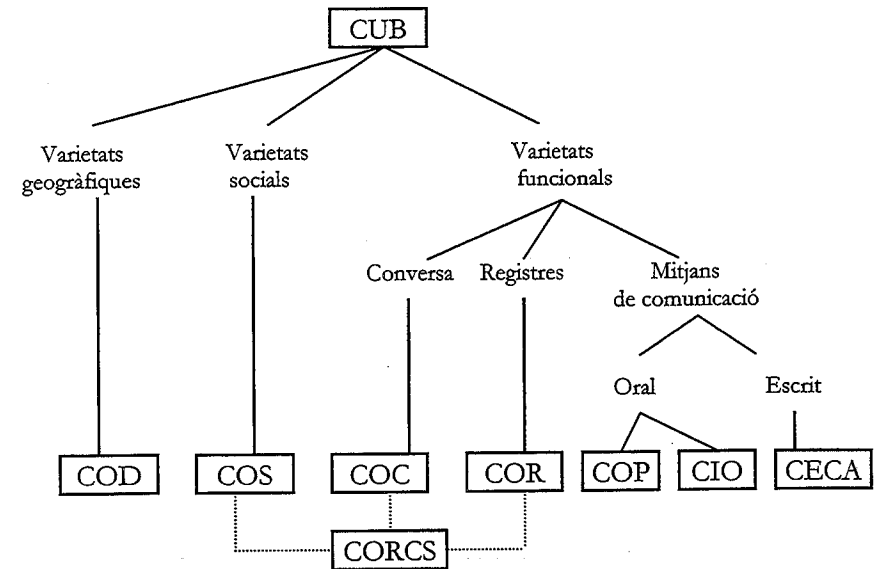


Figura 1. Estructura del CUB

El cos central del CUB el constitueixen els subcorpus orals de varietats geogràfiques (COD), socials (COS), conversa col·loquial (COC), registres (COR) i publicitat (COP). Un bloc a part el formen els subcorpus funcionals CIO (oral) i CECA (escrit), amb unes característiques molt específiques i que, a més, han servit de material exploratori i de referència per al processament de les dades orals.

Els subcorpus de registres (COR), conversa col·loquial (COC) i varietats socials (COS) comparteixen l'objectiu de recollir textos orals de dialectes orientals (fonamentalment centrals). Tots tres constitueixen un bloc específic dins del CUB que es caracteritza per un tractament unificat dels materials: el Corpus Oral de Registres, Conversa Col·loquial i Varietats Socials (CORCS) (veg. 2.2).

<sup>2</sup> L'annex 1 recull els noms dels professors, becaris i col·laboradors voluntaris del CUB des de 1991, i els noms dels estudiants que han aportat materials al Corpus Oral de Conversa Col·loquial.

<sup>3</sup> El CUB ha rebut diversos ajuts econòmics de la CIRIT i de la DGICYT a través del projecte de recerca *La variació en el llenguatge. El català actual* (PB90-0505, PB94-0951, PB97-0889, CS93-1017), del grup de recerca consolidat GEV (SGR95-292, SGR96-104, SGR98-041) i del Centre de Referència en Enginyeria Lingüística de Catalunya. A més, ha rebut ajuts del Vicerectorat de Recerca de la Universitat de Barcelona, destinats especialment a les necessitats de gestió i manteniment de les eines informàtiques.

Descrivim breument, tot seguit, les característiques de cadascun dels subcorpus del CUB.<sup>4</sup>

- **Corpus Escrit del Català Actual (CECA)**  
Conté 7000 pàgines de text procedent de materials de premsa escrita (125 exemplars del diari *Avui*, de Barcelona) que han estat marcats, lematitzats i adaptats als formats estàndard europeus (norma EAGLES, format SGML).
- **Corpus d'Informatius Orals (CIO)**  
Aquest subcorpus està constituït per 240 minuts de programes informatius de ràdio, que han estat digitalitzats, segmentats i transcrits fonèticament.
- **Corpus Oral de Publicitat (COP)**  
El subcorpus de publicitat conté 443 anuncis publicitaris de televisió (423) i ràdio (20) que van ser enregistrats entre 1992 i 1995. Aquests materials han estat parcialment transliterats, codificats i informatitzats.
- **Corpus Oral de Registres (COR)**  
El subcorpus de registres s'ha concebut com un corpus obert al qual es van incorporant mostres de llengua oral representatives de diversos gèneres de parla característics de l'entorn social de l'àrea metropolitana de Barcelona.
- **Corpus Oral de Conversa Col·loquial (COC)**  
Conté 50 mostres de conversa col·loquial, cadascuna de les quals té una durada aproximada de 30 minuts. 10 d'aquestes mostres estan disponibles per a la seva consulta, un cop superat el procés d'homogeneïtzació, revisió i correcció.

<sup>4</sup> Per a més informació sobre els orígens i les característiques del CUB, veg. Payrató, L., E. Boix, M. R. Lloret i M. Lorente (eds.) (1996): *Corpus, corpora*, Barcelona: PPU – Universitat de Barcelona.

- **Corpus Oral de Varietats Socioculturals (COS)**  
Aquest subcorpus conté dades representatives de les varietats socials de l'àrea metropolitana de Barcelona. La mostra, que es troba en la fase constitució, consistirà en 79 entrevistes semidirigides, per a la realització de les quals s'ha adaptat el mètode etnogràfic de la història de vida.
- **Corpus Oral Dialectal (COD)**  
Recull de dades morfològiques i textuais que s'han obtingut mitjançant 490 enquestes dialectològiques realitzades arreu del domini lingüístic català. Inclou les varietats geogràfiques següents: català septentrional, català central, català nord-occidental, balear, valencià i alguerès.<sup>5</sup>

## 2.2 El Corpus Oral de Registres, Conversa Col·loquial i Varietats Socials (CORCS)

### 2.2.1 Introducció

El Corpus Oral de Registres, Conversa Col·loquial i Varietats Socials té com a objectiu l'emmagatzematge de dades de llengua oral (català central contemporani, bàsicament de la Regió de Barcelona), que puguin ser utilitzades en les investigacions sobre la variació funcional (COC i COR) i la social (COS).

El corpus que ens ocupa posa a l'abast de qualsevol investigador un gruix important de materials informatitzats que pot facilitar aquest tipus de recerca, ja que la tasca d'obtenir dades orals era un escull que impedia o, en el millor dels casos, alentia el procés de treball en aquest camp.

CORCS està constituït per tres subcorpus: el Corpus Oral de Registres (COR), el Corpus Oral de Conversa Col·loquial (COC) i el Corpus Oral de Varietats Socials (COS). Es disposa de dues bases de dades actualitzades per a cada subcorpus, que contenen tota la informació disponible dels textos enregistrats i dels participants, respectivament. D'aquesta manera,

<sup>5</sup> Amb aquestes dades s'ha dissenyat una base de dades polivalent que conté entrades relatives a aspectes fonètics, adjectivals (en relació amb el gènere), pronominals (possessius, demostratius, pronoms personals, pronoms àtons, articles i locatius) i relatius a la morfologia verbal regular.

s'hi poden consultar, per exemple, des de la durada de l'enregistrament fins al nombre de participants. CORCS aplega també un conjunt de materials addicionals (alguns en suport audiovisual) que encara no han estat transcrits.<sup>6</sup>

El disseny de cada subcorpus és autònom, ja que els objectius, el procediment de recollida de dades i les dimensions de les mostres són diferents en cadascun. Tanmateix, com que la possible explotació conjunta exigeix una coordinació, es treballa amb les mateixes normes de transcripció del discurs oral (veg. annex 2), basades en tres unitats: el torn de parla, el grup tonal o unitat entonativa i la paraula.

Les principals línies de recerca del CORCS són la descripció de la llengua oral, la comparació de la llengua oral amb l'escrita, l'estudi de les diferències entre les diverses varietats funcionals, socials o culturals i la descripció dels gèneres orals, a més de la investigació quantitativa de lèxic, el funcionament morfosintàctic, els fenòmens de contacte de llengües o l'estructura de la interacció en els fets de parla orals.

Pel que fa als possibles camps d'aplicació, cal destacar l'ensenyament del català com a primera o segona llengua, les tasques de prescripció normativa i la investigació en les diverses indústries de la llengua (reconeixement de la veu, traducció automàtica, intel·ligència artificial, lexicografia, etc.).

Apuntem, tot seguit, les característiques essencials del COR i del COS, subcorpus en procés d'elaboració que, de moment, només es poden consultar parcialment. La descripció detallada del COC s'exposa a l'apartat 3.

### 2.2.2 El Corpus Oral de Registres (COR)

Aquest subcorpus té com a objectiu la recollida d'una selecció de mostres d'esdeveniments o fets de parla prototípics de la comunitat catalanoparlant, i està concebut com un corpus obert (monitor). Els primers 16 textos, que constituïen una mostra pilot, van ser recopilats el 1992-1993 a partir d'una graella que permetia distingir-los segons el camp (genèric o tècnic), el mode (espontani o planificat), el tenor (informatiu o

<sup>6</sup> Es disposa d'un centenar d'entrevistes complementàries, de 45 minuts cadascuna, que provenen del programa de TV3 *Ciudadans*, dirigit per J. Úbeda i F. Escibano. A més, el corpus també conté 10 classes d'educació secundària.

interactiu) i el to (formal o informal).<sup>7</sup> Aquest disseny ha estat posteriorment completat tenint en compte les nocions d'àmbit d'ús i de gènere textual.

En aquests moments, el COR conté 32 interaccions digitalitzades i enregistrades en discos compactes que s'han transcrit ortogràficament i s'han revisat parcialment. A més, es disposa, per a cada mostra, de cinc minuts transcrits fonèticament i d'informacions complementàries de tipus etnogràfic i social.

El COR està pensat per facilitar l'estudi de les tipologies textuales (es disposa de textos narratius, descriptius, argumentatius, expositius, explicatius), les variables de la variació contextual (camp, mode, tenor i to) i el gènere o el fet de parla de què es tracti (conferència, tertúlia, sermó, conte, recepta de cuina, etc.).

### 2.2.3 El Corpus Oral de Varietats Socials (COS)

L'objectiu d'aquest subcorpus és proporcionar un material adequat per a la descripció i l'estudi de les varietats socials del català de l'àrea metropolitana de Barcelona. Amb aquesta finalitat, s'està configurant una mostra de 78 entrevistes semidirigides de final obert per a la realització de les quals s'ha adaptat el mètode antropològic de la història de vida. Aquesta mostra pretén ser representativa de la comunitat de parla catalana des d'un punt de vista social i cultural. S'ha confeccionat a partir dels criteris de selecció dels parlants de l'*Enquesta de la Regió Metropolitana de Barcelona* (Subirats *et al.*, 1992),<sup>8</sup> en què es distingeixen diferents grups socials:

- Elits, alta burgesia
- Noves classes mitjanes
  - a) Tècnics alts, professionals
  - b) Empleats, tècnics mitjans
- Velles classes mitjanes
  - a) Petits empresaris

<sup>7</sup> Veg. Payrató, L. (ed.) (1998): *Oralment. Estudis de variació funcional*, Publicacions de l'Abadia de Montserrat - Universitat de Barcelona, Barcelona.

<sup>8</sup> Subirats, M., M. Masats i C. Sánchez (1992): *Enquesta de la Regió Metropolitana de Barcelona 1990, vol. 4 de l'Informe General de l'estudi. Educació, llengua i hàbits culturals*, Institut d'Estudis Metropolitans de Barcelona, Bellaterra.

- b) Comerciants autònoms
- Sectors treballadors
  - a) Contramestres, capatassos
  - b) Treballadors manuals
  - c) Treballadors manuals vells i aturats en edat de treballar
- Lumpen / marginals

Pel que fa a l'estat actual de les dades, es disposa de 20 interaccions transcrits ortogràficament, revisades, digitalitzades i enregistrades en discos compactes, cadascuna de les quals dura entre 30 i 75 minuts. A més, la informació recollida sobre els parlants inclou les seves característiques socioculturals fonamentals (edat, sexe, nivell d'estudis i ocupació) i d'altres de complementàries.

### 3 El Corpus Oral de Conversa Col·loquial

#### 3.1 Objectius i mostres

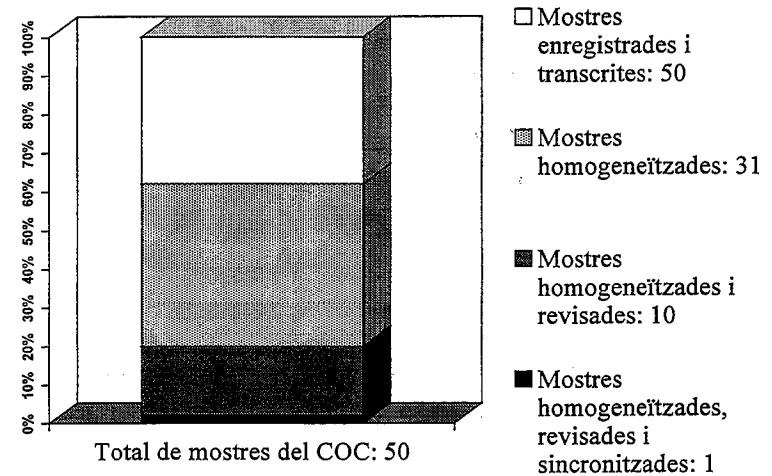
L'objectiu principal del Corpus Oral de Conversa Col·loquial (COC) és subministrar als investigadors mostres de llengua catalana extretes de la conversa col·loquial, entesa com a gènere prototípic de les situacions de parla dialogades de to informal, tenor interactiu, camp genèric i mode oral espontani. El propòsit, doncs, és oferir materials de llengua corrent per desenvolupar estudis de descripció i anàlisi del català col·loquial des d'una perspectiva bàsicament pragmàtica (anàlisi conversacional, fenòmens textuais i gramaticals, etc.).

A més de proporcionar fonts d'estudi per a la recerca, el COC també pot ser una eina molt útil en el camp de la didàctica. Els materials que aquí es presenten, en efecte, posen a disposició de mestres i professors mostres de català col·loquial que poden ser de gran utilitat en l'ensenyament del català com a segona llengua, en l'anàlisi comparativa de registres, en la formació de transcriptors, etc.

El COC conté 50 mostres d'una durada aproximada de 30 minuts cadascuna, que fan un total de 25 hores d'enregistrament. Totes les mostres s'han transcrit i posteriorment s'han digitalitzat i gravat en un total de 8 discos compactes. La majoria, concretament 31 mostres, també s'han homogeneïtzat, de les quals 10 han estat revisades i corregides. Aquest

conjunt de 10 mostres constitueix el bloc principal del corpus (anomenat Central Fonamental). Finalment, 1 d'aquestes 10 mostres també ha estat completament sincronitzada.<sup>9</sup>

L'estat actual de les dades pel que fa al procés de tractament, explicat amb més detall en l'apartat 3.3, es resumeix en percentatges en el gràfic 1.



Gràfic 1. Estat actual del COC amb relació al procés de tractament de les dades

Del conjunt de les 50 converses, hi ha constància que 24 han estat enregistrades de manera subreptícia i 6 de manera no subreptícia. Són converses amb una mitjana de 4 parlants (3 com a mínim i 11 com a màxim). La majoria dels participants del COC tenen el català oriental central com a varietat pròpia i aproximadament un 70% tenen una edat compresa entre els 20 i els 40 anys. La totalitat dels parlants són de classe social mitjana (noves i velles classes mitjanes) i de sectors treballadors; (veg. 2.2.3).

<sup>9</sup> Vegen l'apartat 3.3.

### 3.2 Classificació de les mostres

Les mostres s'han classificat en cinc grans blocs segons les característiques que s'especifiquen a continuació:

- Central Fonamental

Mostres en què la varietat dialectal dels parlants és el català oriental central i que compleixen la totalitat dels criteris següents:

a) Criteris respecte a la situació comunicativa:

Mostres de to informal, tenor interactiu i no transaccional i pertanyents a una situació familiar amical (sopars o dinars familiars, tertúlies de sobretaula, trobades d'amics, etc.).

b) Criteris respecte a les característiques dels parlants:

Parlants adults (majors de 18 anys) i sense formació especialitzada en llengua catalana.

c) Criteris respecte a les característiques dels enregistraments:

Mostres d'una durada d'entre 25 i 35 minuts aproximadament, enregistrades de manera subreptícia i amb una qualitat de so prou bona.

d) Criteris respecte a les informacions complementàries:

Mostres amb informació pertinent sobre les característiques socioculturals dels parlants i la situació comunicativa, i amb el permís escrit de l'autor i dels participants per a la utilització dels enregistraments.

Total de mostres: 10

- Central Complementari

Mostres en què la varietat dominant és el català oriental central, però que no compleixen la totalitat dels criteris esmentats en el primer apartat.

Total de mostres: 15

- No Central Fonamental

Mostres en què la varietat dominant no és el català oriental central, però que compleixen els criteris de selecció del primer apartat.

Total de mostres: 10

- No Central Complementari

Mostres en què la varietat dominant no és el català oriental central i que no compleixen la totalitat dels criteris de selecció.

Total de mostres: 4

- Altres

Mostres en què algun dels parlants té el castellà com a llengua de procedència familiar o com a llengua d'ús més habitual, i que poden no complir la totalitat dels criteris de selecció.

Total de mostres: 11

Convé destacar que les 10 mostres que pertanyen al bloc Central Fonamental són les que ja han passat totes les fases de la revisió i, per tant, les que a hores d'ara ofereixen una transcripció més acurada. En l'apartat 3.4 s'especifiquen més detalladament les característiques de cadascuna de les mostres que pertanyen a aquest bloc.

### 3.3 Recollida i tractament de les dades

El procés de recollida i tractament consta de cinc fases:

- Enregistrament:

L'observador participant enregistra subreptíciament la interacció amb un magnetòfon convencional. Posteriorment la gravació es digitalitza mitjançant el programa de tractament de veu *SoundEdit 16* (versió 2.1) i s'enregistra en discos compactes amb el programa *Adaptec Toast*.

- Recopilació d'informació:

Es creen bases de dades amb la informació dels participants (edat, sexe, ocupació, llengua d'origen familiar, llengua de domini actual...) i altres informacions generals que fan referència als intercanvis comunicatius per tal de facilitar les cerques (codi d'identificació, data, durada, nombre de mots...).

- **Transcripció i homogeneïtzació:**  
L'autor de l'enregistrament realitza una primera transcripció ortogràfica que després es repassa perquè s'adeqüi als criteris de presentació formal convinguts i s'evitin així problemes d'homogeneïtat.
- **Revisió:**  
Es converteixen les transcripcions ortogràfiques en bases de dades per contrastar-les amb la veu digitalitzada i revisar-les (veg. annex 3 i annex 4.1).
- **Sincronització:**  
Està previst sincronitzar la veu amb les transcripcions en bases de dades mitjançant el programa *Paral·lelcial* creat per Lluís de Yzaguirre (veg. annex 4.2). Com s'ha dit més amunt, fins ara només una conversa ha estat completament sincronitzada.

### 3.4 Descripció del bloc Central Fonamental

#### 3.4.1 Característiques generals

El bloc Central Fonamental conté 10 enregistraments de conversa col·loquial (tenor interactiu, to informal, canal oral espontani i tema general) en una situació familiar o amical realitzats durant els anys 1992, 1993, 1994 i 1997:

Mostra	Situació	Data
01	sopar al menjador	1992-93
02	xerrada al sofà després d'un dinar a fora de casa	14/2/93
03	dinar al menjador	DG/5/93
04	dinar	19/2/93
05	xerrada al menjador després d'un partit de futbol	nit/8/1997
06	sopar i cafè al menjador	1992-93
07	abans de dinar al menjador	10/11/93
08	sobretaula al sofà amb la televisió engegada	1993-94
09	sobretaula al menjador	8/12/93
10	visita de cortesia al menjador	8/12/93

Taula 1. Mostra, situació i data de l'enregistrament

Aquest corpus consta d'un total aproximat de 281,67 minuts (28,16 minuts de mitjana per cada text o enregistrament) i 70493 mots (7049,3 mots de mitjana), comptant-hi les inicials dels parlants i tots els símbols establerts en el sistema de transcripció. Com es pot observar a la taula 2, cada enregistrament té una durada d'un mínim de 23 minuts i un màxim de 31, i conté un mínim de 5200 mots i un màxim de 8966.

Mostra	Minuts	Mots
01	30,50	6445
02	28,05	6603
03	29,07	8966
04	23,07	5781
05	23,15	6765
06	29,06	5266
07	30,23	8613
08	31,09	6632
09	30,27	7228
10	27,18	8194
<b>TOTAL</b>	<b>281,67</b>	<b>70493</b>

Taula 2. Especificació de la durada i del nombre de mots

La qualitat de so dels enregistraments no és gaire bona a causa de l'elevat nombre de participants, dels mitjans tècnics utilitzats (una gravadora convencional) i, com ja s'ha especificat a l'apartat 3.3, del seu caràcter subreptici:

Mostra	Participants	Qualitat de so
01	5	regular
02	5	bona
03	8	regular
04	7	regular
05	4	excel·lent
06	4	bona
07	3	bona
08	3	regular
09	11	regular
10	7	regular

Taula 3. Nombre de participants i qualitat de so de les mostres

Un altre tret general dels locutors d'aquesta part del corpus, com s'ha apuntat a l'apartat 3.2, és que es tracta de parlants adults de dialectes centrals (sobretot de la regió de Barcelona) que, en general, no tenen coneixements especialitzats de català.

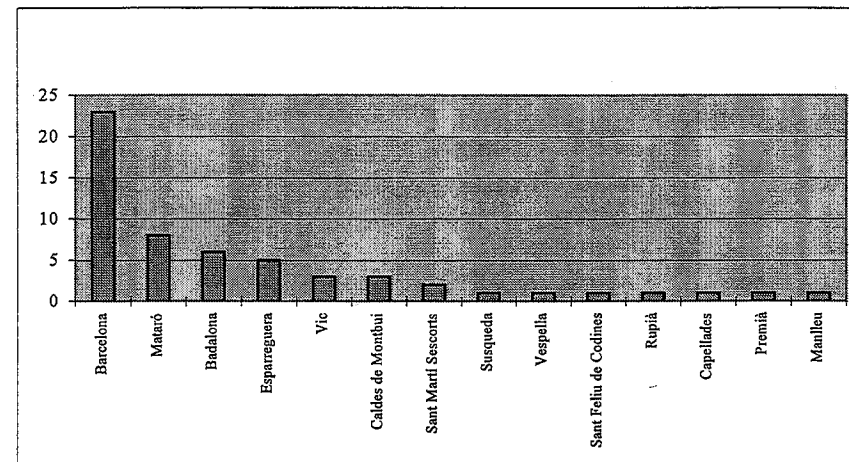
### 3.4.2 Característiques dels parlants

Exposem en els dos subapartats següents les característiques principals dels parlants del bloc Central Fonamental del Corpus Oral de Conversa Col·loquial pel que fa a la seva distribució geogràfica i social.<sup>10</sup>

#### 3.4.2.1 Distribució geogràfica

Hi ha 23 persones de Barcelona capital i 34 de fora d'aquesta ciutat, fet que representa un 40,35% de Barcelona i un 59,64% de la resta: 23 parlants de Barcelona (Barcelonès), 8 de Mataró (Maresme), 6 de Badalona (Barcelonès), 5 d'Esparreguera (Baix Llobregat), 3 de Vic (Osona), 3 de Caldes de Montbui (Vallès Occidental), 2 de Sant Martí Sescorts (Osona), 1 de Susqueda (La Selva), 1 de Vespella (Osona), 1 de Sant Feliu de Codines (Vallès Oriental), 1 de Rupió (Baix Empordà), 1 de Capellades (Anoia), 1 de Premià (Maresme) i 1 de Manlleu (Osona).

Aquesta distribució queda reflectida en el gràfic 2.

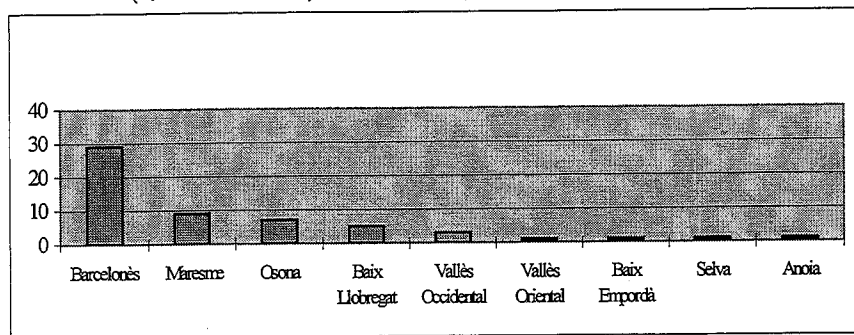


<sup>10</sup> A l'annex 5 s'ofereix una taula amb informació individualitzada dels parlants d'aquest bloc del corpus.



Gràfic 2. Distribució geogràfica dels parlants per poblacions

Per comarques, hi ha 29 persones del Barcelonès (50,87% del total), 9 del Maresme (15,78% del total), 7 d'Osona (12,28% del total), 5 del Baix Llobregat (8,77% del total), 3 del Vallès Occidental (5,26% del total), 1 del Vallès Oriental (1,75% del total), 1 del Baix Empordà (1,75% del total), 1 de la Selva (1,75% del total) i 1 d'Anoia (1,75% del total).



Gràfic 3. Distribució geogràfica dels parlants per comarques

Com es pot observar en el gràfic 3, en aquesta part del COC hi ha un predomini clar dels parlants del Barcelonès enfront dels de les altres comarques, la major part de la Regió de Barcelona:<sup>11</sup> el Maresme, el Baix Llobregat, el Vallès Occidental i el Vallès Oriental.

El 82,45% del conjunt de persones d'aquest bloc (47 parlants), doncs, procedeixen de la Regió de Barcelona, i el 17,54% restant (10 individus) prové de comarques de fora d'aquesta regió, concretament d'Osona, el Baix Empordà, la Selva i Anoia.

### 3.4.2.2 Distribució social

Dels 57 individus que componen el bloc Central Fonamental del COC, 35 (el 61,4%) són dones i 22 (el 38,5%) són homes.

<sup>11</sup> Segons la GEC (1971), la Regió de Barcelona consta del Barcelonès, el Baix Llobregat, l'Alt Penedès, el Garraf, el Baix Penedès, el Vallès Occidental, el Vallès Oriental i el Maresme.

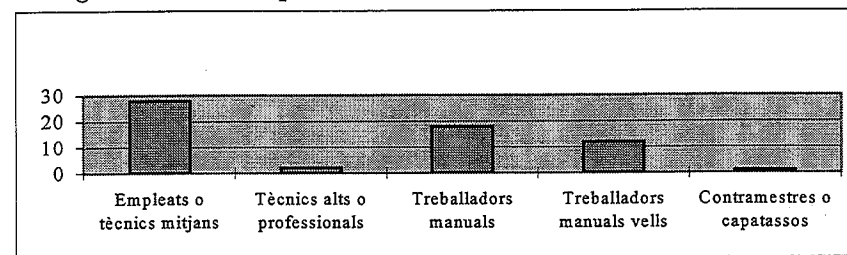
Pel que fa a l'estratificació social dels parlants, s'estableixen dos grups:<sup>12</sup>

a) 30 parlants de noves classes mitjanes (52,63%), 28 dels quals són empleats o tècnics mitjans (49,12%) i 2 tècnics alts o professionals (3,5%).

b) 25 parlants de sectors treballadors (43,85%), 18 dels quals són treballadors manuals (31,57%), 6 treballadors manuals vells (10,52%) i 1 contramestre o capatàs (1,75%).

Cal afegir, finalment, que no s'ha estipulat grup social en el cas de 2 nens (3,5%) per manca de dades i perquè la seva participació lingüística o comunicativa en les converses és mínima.

El gràfic 4 il·lustra aquesta distribució.



Gràfic 4. Distribució social dels parlants

### 3.5 Consulta i utilització de les dades

La consulta d'aquests materials,<sup>13</sup> com s'ha apuntat anteriorment, està oberta a tot investigador que hi estigui interessat per a la realització de tesis, projectes d'investigació, articles, etc., finalitat per a la qual ja s'han utilitzat en alguns treballs (vegeu-ne la relació a l'annex 6).

Amb aquest objectiu, s'estableix un contracte d'intercanvi pel qual l'investigador en qüestió es compromet a fer constar l'origen de les dades consultades (el Corpus de Català Contemporani de la Universitat de

<sup>12</sup> Es parteix de la proposta de l'Enquesta de la Regió Metropolitana de Barcelona (veg. apartat 2.2.3., en especial la nota 8).

<sup>13</sup> Està previst que aquests materials estiguin disponibles en discos compactes que inclouin la veu digitalitzada en format *wave*, les transcripcions ortogràfiques de cada conversa en els formats RTF i MicrosoftWord 6.0 de PC, una base de dades dels informants en MicrosoftFoxPro i MicrosoftAccess, i aquest mateix article.

Barcelona) i les fonts de finançament del projecte en què els materials s'inscriuen (PB90-0505, PB94-0951 i PB97-0889 de la DGICYT i CS93-1017, SGR95-292, SGR96-104 i SGR98-041 de la CIRIT). A més, es compromet a donar al Departament de Filologia Catalana un exemplar de qualsevol publicació o treball que en resulti.

La sol·licitud de consulta<sup>14</sup> pot dirigir-se a l'adreça següent:

CORPUS ORAL DE CONVERSA COL·LOQUIAL  
 Departament de Filologia Catalana (Secció de Lingüística)  
 Universitat de Barcelona  
 Gran Via de les Corts Catalanes, 585  
 E-08007 Barcelona

## Annexos

### *Annex 1. Investigadors*

Des de 1991, han format part de l'equip de recerca del CUB els investigadors següents (els membres de l'equip actual apareixen en lletra cursiva):

- Professors universitaris  
*Núria Alturo*, Antoni M. Badia i Margarit (investigador principal des de 1991 fins a 1995), *Josep Besa*, *Emili Boix*, *Sebastià Bonet*, M. Teresa Cabré (UB, UPF), *Josefina Carrera*, *Esteve Clua* (UPF), *M. Rosa Lloret*, Mercè Lorente (UB, UPF), *Mar Massanell*, M. Àngels Massip, Josep Moran, Anna Oller, *Lluís Payrató*, *M. Pilar Perea*, Lúdia Pons, Joan Solà, *Amadeu Viana* (UDL), *Joaquim Viaplana* (investigador principal des de 1995), *F. Xavier Vila* i Lluís de Yzaguirre (UB, UPF).
- Becaris predoctorals (CIRIT i DGICYT)  
*Òscar Bladas*, Ignasi Clemente, *Mar Massanell*, *Marta Payà*, *M. Pilar Perea*, *Clàudia Pons*, *Marta Torres* i *F. Xavier Vila*.

<sup>14</sup> També es pot sol·licitar informació a les adreces de correu electrònic de Núria Alturo, Òscar Bladas, Anna Oller, Marta Payà, Lluís Payrató o Marta Torres (*alturo@lincat.ub.es*, *bladas@lincat.ub.es*, *oller@lincat.ub.es*, *paya@lincat.ub.es*, *payrato@lincat.ub.es*, *torres@lincat.ub.es*).

- Becaris i col·laboradors voluntaris de segon i tercer cicle de la UB i col·laboradors externs  
*Rosa Aponte*, Josep Azemà, *Laura Babot*, Antoni Bernadó (UDL), *Judit Calvo*, *Maria Carbó*, *Maria Casals*, *Eva Casasús*, *Meritxell Coll*, *Antoni Corcoll*, *Elisabet Curt*, *Mireia Galindo*, *Marta Grané*, *David Graupera*, *Mar Massanell* (IEC), *Amparo Miralles*, *Montserrat Montraveta*, *Anna Oller*, *Marta Payà*, *M. Pilar Perea*, *Clàudia Pons*, *Maria Prieto*, *Berta Puig*, *Berta Pujol*, *Oriol Ramon*, *Carles de Rosselló*, *Trinitat Ruiz*, *Xavier Rull*, *Albert Soler* (UB, UAB) i *Eugènia Ustrell*.
- Col·laboradors responsables de la recollida i la transcripció inicial de les dades del COC<sup>15</sup>  
*V. Alzina*, *M. Artigas*, *C. Bach*, *M. Calafell*, *J. Calvo*, *G. Camps*, *M. Carnicero*, *E. Casasús*, *À. Castells*, *J. Cervera*, *M. P. Coll*, *G. Coma*, *N. Coma*, *N. Comas*, *F. Dedeu*, *S. Domínguez*, *F. X. Fargas*, *M. Font*, *E. Garriga*, *N. Guasch*, *A. Guiu*, *M. Lacorte*, *A. Llombart*, *N. Madaula*, *J. Marín*, *C. Miralles*, *M. Moliné*, *T. Muñoz*, *I. Obradós*, *M. Olivella*, *A. Oller*, *P. Piqué*, *B. Puig*, *C. Rusiñol*, *S. Sardà*, *R. Saurí*, *G. Senabre*, *M. Sánchez*, *M. Soler*, *I. Tàpies*, *M. Thorson*, *D. Torner*, *E. Vergés*, *R. Vidal*, *L. Vila*, *C. Vilaró*, *Y. Villanueva*.

<sup>15</sup> Citem el nom de tots els estudiants que han posat a disposició del COC els enregistraments i transcripcions de converses que van realitzar com a part del seu treball de curs. Cal dir, però, que no tots aquests materials han estat admesos: s'han rebutjat els enregistraments que no reuneixen les característiques necessàries exigides per formar part del COC (bona qualitat de so, caràcter subreptici de l'enregistrament, procedència dialectal oriental dels parlants, etc.; veg. 3.2).

Annex 2. Convencions de transcripció<sup>16</sup>

## 2.1 Aspectes prosòdics

Seqüència tonal terminal	
descendent	\
ascendent	/
de manteniment	—
Grup tonal truncat	--
Manteniment de l'entonació	
ascendent	{(A) text afectat}
descendent	{(B) text afectat}
Estrafer la veu	{(EV) text afectat}
Intensitat	
força	{(F) text afectat}
molt forta	{(FF) text afectat}
fluixa	{(P) text afectat}
molt fluixa	{(PP) text afectat}
Temps	
accelerat	{(AC) text afectat}
desaccelerat	{(DC) text afectat}
Allargament breu, mitjà i llarg	: :: :::

## 2.2 Aspectes vocals

Riure simultani amb el discurs	{(@) text afectat}
Riure no simultani amb el discurs (un símbol per síl·laba)	@

<sup>16</sup> Veg. més referències sobre les convencions i el sistema de transcripció a:  
 Edwards, J.A. i M. D. Lampert (eds.) (1993): *Talking data: Transcription and coding in discourse research*. Londres: Erlbaum.  
 Payrató, L. (1995): «Transcripción del discurso coloquial». Dins Cortés, L. (ed.): *El español coloquial*, Almería: Universidad de Almería, 45-70. Trad. cat.: «Transcripció del discurs oral». Dins Payrató, L., E. Boix, M. R. Lloret i M. Lorente (eds.): *Corpus, corpora*, PPU – Universitat de Barcelona, Barcelona, 1996.  
 Payrató, L. (ed.) (1998): *Oralment. Estudis de variació funcional*, Publicacions de l'Abadia de Montserrat – Universitat de Barcelona, Barcelona.

Fragment llarg amb la durada especificada	@R(durada)R@
Inhalació i exhalació	(INH) (EXH)

## 2.3 Pauses i encavalcaments

Pausa	
molt breu ( $0.1 < p < 0.3$ )	(. durada)
breu ( $p < 1$ )	(.. durada)
mitjana ( $1 \leq p < 3$ )	(... durada)
llarga ( $p \geq 3$ )	(... durada)
Encavalcament	[text afectat]

## 2.4 Fragments conflictius

Dubte sobre el text transcrit	{(??) text afectat}
Fragment intel·ligible	
Un signe per síl·laba	x
Fragment llarg amb la durada especificada	xX(durada)Xx

## 2.5 Altres aspectes

Segona llengua	{(L2) text afectat}
Truncament de mot	-
Elisió	ús de parèntesis per marcar el so elidit
Comentari del transcriptor	(comentari)
Descripció de fenòmens (p.ex. sorolls, gestos, etc.)	(fenomen)

## Annex 3. Exemple de transcripció

1 PPP: (. 0.55) {(??) a més} vaig mort jo eh\  
 2 (. 0.41) goita\  
 3 (. 0.47) avui no has sentit el despertador\  
 4 i jo tampoc el sentia eh/  
 5 (. 0.24) al final crec que ha [toca- --]  
 6 MMM: [no\  
 7 eh que poder no [te l'he sent- --]  
 8 PPP: [ha tocat] tres vegades\  
 9 MMM: (... 1.44) vols dir que no l'he sentit i t'he do[nat un cop]/  
 10 PPP: [no::\  
 11 no\  
 12 no ho cr[ec\  
 13 no\  
 14 MMM: [ah\  
 15 jo diria que sí\  
 16 PPP: (. 0.41) no t'he pas sentit jo a tu\  
 17 MMM: ah\  
 18 doncs jo t'he donat un cop [Joan\  
 19 PPP: [xxx--]  
 20 si dormies com una lirona\  
 21 MMM: (. 0.35) oh\_  
 22 perquè m'he posat a dormir\_  
 23 p(e)rò\_  
 24 @  
 25 dormies com una lirona\  
 26 sí\  
 27 perquè m'he posat a dormir p(e)rò [t'he--]  
 28 PPP: [està] bé aquesta paraula\  
 29 CME: (. 0.27) eh/  
 30 PPP: (. 0.36) si està ben dita\_  
 31 això de lirona\  
 32 MMM: és igual\  
 33 CME: no sé ni què vol dir\  
 34 PPP: (. 0.60) no:/  
 35 mira-ho [al diccionari\  
 36 MMM: [no saps què vol dir l]irona\  
 37 que és @@ (. 0.15) {( @ ) una miqueta fluixa } @@  
 38 p- --  
 39 PPP: no::\  
 40 jo no volia pas dir això\  
 41 MMM: (. 0.23) ah\  
 42 lirona vol dir ai[xò\  
 43 PPP: [oh:]

44 bueno p(e)rò jo no ho [volia dir això doncs\  
 45 MMM: [que no s'està gaire] bé\  
 46 PPP: (. 0.17) al tanto eh\_  
 47 MMM: (. 0.32) p(e)rò::\_  
 48 jo t'he donat un cop aixís\  
 49 (. 0.41) jo sí que l'he sen[tit 1]  
 50 PPP: [1 jo] [2 no me'n recordo\  
 51 MMM: [2 i t'he donat un cop] aixís (. 0.74) [3 jo\  
 52 ((MMM torna a fer un senyal de cop amb el colze))  
 53 PPP: [3 no] me'n recordo que m'hagis donat cap cop\  
 54 MMM: sí\  
 55 ja\  
 56 perquè tu dormies\  
 57 (. 0.66) i jo he pe(n)sat\_  
 58 ja són--  
 59 ja: de--  
 60 això sí\  
 61 perquè com m'he desper[tat--]  
 62 PPP: [te]nia una so:n\_  
 63 MMM: com m'he despertat hi he pensat en això\  
 64 Joan\  
 65 (. 0.30) t'he donat un cop\  
 66 [t'he xx de sota de--]  
 67 PPP: [jo m'he despertat--]  
 68 en Jaume se n'ha anat a dormir a do- --  
 69 casi a dos quarts de tres eh\  
 70 CME: (. 0.37) escrivint a màquina\  
 71 PPP: (. 0.22) no ho sé\  
 72 (. 0.10) [{(??) potser sí\  
 73 MMM: [fins a--]  
 74 jo jo era al llit que encara escrivia\  
 75 PPP: escrivia/  
 76 [no l'he sentida jo\  
 77 MMM: [sí\  
 78 sort] d'això\  
 79 PPP: (... 1.13) {(??) xx la senyora Lola\  
 80 [oi/]}  
 81 MMM: [p(e)rò] si em va dir que només tenia que fer dugues o tres pàgines [d'allò\  
 82 CME: [m'ha dit] que és un treball\  
 83 (. 0.23) i encara està fent la tapa i tot\  
 84 PPP: (. 0.80) ves\  
 85 m'ha despertat quan baixava la la a: el llit\  
 86 (. 0.76) br[u::m\  
 87 MMM: [xu:c\  
 88 sí:

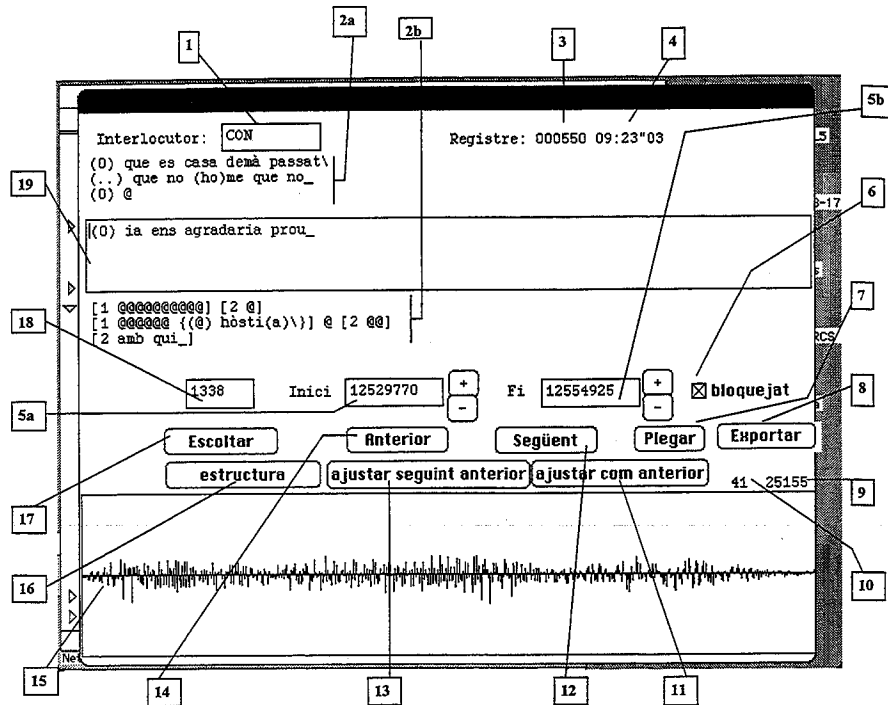
89 jo també ho he sentit\  
 90 PPP: cony\  
 91 què és això\  
 92 CME: (. 0.13) doncs [jo he dormit\_  
 93 PPP: [llavons sí que-]  
 94 CME: (.. 0.43) oh::\  
 95 [xxx]  
 96 PPP: [jo sí] que m'he despertat llavons\  
 97 amb això del llit\  
 98 (.. 0.78) i llavons no podia dormir\  
 99 (.. 0.38) cagondena\  
 100 i llavons tinc els peus que em [cremen\  
 101 CME: ((a MMM)) [vagi a mirar] l'arròs\  
 102 perquè em sembla [que\_] ((MMM se'n va a la cuina)  
 103 PPP: [te]nia que treure els peus a fora\_  
 104 i llavòrens tenia tos\_  
 105 MMM: (. 0.15) ((a CME)) ja\  
 106 marxa l'aigua de l'arròs\  
 107 PPP: (. 0.26) haig de treure el peus a fora sempre\  
 108 cada nit\  
 109 (... 1.06) fora del llit\  
 110 CME: (.. 0.62) @@@[@]  
 111 PPP: [ca]gon coi\  
 112 sí\  
 113 no puc dormir de tan calents que són\  
 114 (... 1.05) és de tant tantes hores dret home\  
 115 (... 1.24) ahir vaig estar set hores: seguides dret\  
 116 CME: (. 0.16) ja\  
 117 PPP: (... 1.69) fent neules\  
 118 (... 3.15)  
 119 PPP: ara jo dic\_  
 120 hi hi anireu a futbol o no/  
 121 CME: (..0.46) ah\  
 122 no sé\

## Annex 4. Eines d'ajut per a la transcripció

## 4.1 Mostra de pantalla amb FoxPro i SoundEdit

The screenshot displays the SoundEdit software interface. At the top, the menu bar includes File, Edit, View, Sound, Effects, and Windows. The time is 16:54. Below the menu bar is a table with the following columns: Id, Conté, N\_i, N\_f, Durada, and Mitjana. The table contains 20 rows of transcription data. Below the table, there are controls for Selection (start and end points at 450000 and 500000) and Controls (Record, Play, and other buttons). At the bottom, a waveform is visible, and the status bar shows 8 Bits / 22,050 kHz.

Id	Conté	N_i	N_f	Durada	Mitjana
PPP	[ai que estic cansat\]	202758	225540	1020	51
PPP	ai\	230741	239125	370	123
MMM	i::	243125	261246	810	202
MMM	ha dit_	261246	270305	400	57
MMM	bueno\	272705	281078	370	61
MMM	després se l'ha mirat el metge_	281078	312484	1410	45
MMM	i::	324084	331085	310	103
MMM	abans de posa'ls-hi les gotes als ulls_	331085	369284	1720	44
MMM	el metge_	370284	390000	880	88
MMM	li ha mirat [la.--]	390000	421782	1430	84
CME	[doni'm això] que ho porto a la cuina\ xx xx	408000	438251	1360	37
CME	xx xx	450751	461713	490	98
MMM	el metge li ha mirat la: la graduació_	461713	525025	2850	75
MMM	allò que fan llegir lletres\ i::	525025	552018	1210	43
MMM	i::	562518	569064	290	72

4.2 Mostra de pantalla amb *Paral·lelcial*

1. Caixa d'informe de l'interlocutor.
- 2a.Registres anteriors al text seleccionat.
- 2b.Registres posteriors al text seleccionat.
3. Indicador del número de registre del text.
4. Indicador de la referència temporal.
- 5a.Caixa d'informe d'inici de selecció.
- 5b.Caixa d'informe de fi de selecció.
6. Quadre binari per bloquejar el final de la selecció.
7. Botó per anar al menú principal.
8. Botó que crea un arxiu de so del registre editat.
9. Referència numèrica que indica el rang de mostres que hem representat.

10. Referència numèrica que indica cada quantes mostres es fa una representació aproximada.
11. Botó per ajustar els límits del registre seleccionat amb les del registre anterior.
12. Botó per seleccionar el registre posterior.
13. Botó per ajustar la selecció d'inici del registre editat amb el mateix valor que la selecció de fi de l'anterior.
14. Botó per seleccionar el registre anterior.
15. Oscil·lograma.
16. Botó per canviar alguns aspectes de l'estructura de la base de dades.
17. Botó per escoltar el fragment seleccionat.
18. Caixa d'informe del factor d'increment i de decrement de la selecció d'inici i de fi.
19. Caixa d'informe del text editat.

Annex 5. Característiques específiques dels parlants del bloc Central  
Fonamental del COC

Mostra	Parlant	Edat	S	Població	Ocupació	Grup social
01	DIT	23	d	Barcelona	estudiant	nova classe mitjana, b)
01	LAU	48	d	Barcelona	administrativa	sector treballador, b)
01	MMA	15	d	Barcelona	estudiant	sector treballador, b)
01	REP	48	h	Barcelona	perit químic	nova classe mitjana, a)
01	ROM	22	h	Barcelona	estudiant	nova classe mitjana, b)
02	NAA	26	d	Esparreguera	infermera	nova classe mitjana, b)
02	FAN	26	h	Esparreguera	estudiant	nova classe mitjana, b)
02	HIL	29	h	Esparreguera	enginyer	nova classe mitjana, b)
02	MEL	56	h	Esparreguera	obrer especialitzat	sector treballador, b)
02	SUS	23	d	Esparreguera	estudiant	nova classe mitjana, b)
03	AMM	57	d	Mataró	assistent social	nova classe mitjana, b)
03	ARN	20	h	Mataró	estudiant	nova classe mitjana, b)
03	NAT	14	h	Mataró	estudiant	sector treballador, b)
03	MIA	19	d	Mataró	estudiant	nova classe mitjana, b)
03	MER	48	d	Mataró	professor de secundària	nova classe mitjana, b)
03	RAS	74	d	Mataró	llevadora jubilada	nova classe mitjana, b)
03	OSE	21	d	Mataró	estudiant	nova classe mitjana, b)
03	TON	46	h	Mataró	cap servei manteniment	sector treballador, a)
04	ANA	4	d	Badalona	---	---
04	COC	33	d	Badalona	tècnic turisme	nova classe mitjana, b)
04	JOJ	70	h	Badalona	jubilat	nova classe mitjana, b)
04	MAT	60	d	Badalona	mestressa casa	sector treballador, b)
04	MEC	23	d	Badalona	aturada	nova classe mitjana, b)
04	RON	36	h	Susqueda	professor	nova classe mitjana, b)
04	ONI	30	h	Badalona	entrenador esportiu	sector treballador, b)
05	EUU	23	d	Barcelona	estudiant	nova classe mitjana, b)
05	MJJ	23	h	Barcelona	estudiant	nova classe mitjana, b)
05	PUY	23	h	Barcelona	quiromassatgista	sector treballador, b)
05	TRS	23	d	Barcelona	estudiant	nova classe mitjana, b)
06	ANI	22	d	Barcelona	estudiant	sector treballador, b)
06	BEB	19	d	Premià	estudiant	nova classe mitjana, b)
06	NIA	22	d	Barcelona	estudiant	nova classe mitjana, b)
06	MAM	19	d	Barcelona	estudiant	sector treballador, b)
07	CME	22	d	Barcelona	estudiant	nova classe mitjana, b)
07	MMM	52	d	Capellades	mestressa casa	sector treballador, c)
07	PPP	64	h	Rupià	pastisser	sector treballador, b)

08	FSC	30	h	Caldes de Montbui	estudiant, corrector cat.	nova classe mitjana, b)
08	PEI	67	d	Caldes de Montbui	botiguera	sector treballador, b)
08	RRR	59	d	Caldes de Montbui	mestressa casa	sector treballador, c)
09	AVV	85	d	Barcelona	mestressa casa	sector treballador, c)
09	DDD	65	h	Barcelona	jubilat	sector treballador, c)
09	ELE	24	d	Barcelona	estudiant	sector treballador, b)
09	LIA	25	d	Barcelona	administratiu	sector treballador, b)
09	JJJ	28	h	Barcelona	enginyer	nova classe mitjana, b)
09	IDA	56	d	Sant Feliu de Codines	mestressa casa	sector treballador, c)
09	MRC	3	h	Barcelona	---	---
09	MSA	27	d	Barcelona	documentalista	nova classe mitjana, b)
09	NNN	28	d	Barcelona	correctora	nova classe mitjana, b)
09	OOO	25	h	Barcelona	estudiant	nova classe mitjana, b)
09	SER	54	d	Barcelona	professora	nova classe mitjana, b)
10	ALL	25	d	Vic	advocada	nova classe mitjana, a)
10	GEM	18	d	Vic	oficiala perruqueria	sector treballador, b)
10	JON	59	h	Manlleu	camioner	sector treballador, b)
10	MAG	47	d	S. Martí Sescorts	mestressa casa	sector treballador, c)
10	RIA	52	d	S. Martí Sescorts	obrera	sector treballador, b)
10	SEP	50	h	Vespella	pagès	sector treballador, b)
10	VIE	29	h	Vic	auxiliar d'infermeria	sector treballador, b)

Taula 4. Característiques socioculturals dels parlants (1)

Mostra	Parlant	Estudis	Català	LPF	LDA	LHA	CO
01	DIT	univs.	fil.cat.	C	C	C	SÍ
01	LAU	secunds.	nivell C	C	C	C	NO
01	MMA	secunds.	no	C	C	C	NO
01	REP	univs.	no	C	C	C	NO
01	ROM	univs.	no	C	C	C	NO
02	NAA	univs.	no	C	C	C	NO
02	FAN	univs.	no	C	C	C	NO
02	HIL	univs.	no	C	C	C	NO
02	MEL	prim.	no	C	C	C	NO
02	SUS	univs.	fil.cat.	C	C	C	SÍ
03	AMM	univs.	no	C	C	C	NO
03	ARN	univs.	no	C	C	C	NO
03	NAT	secunds.	no	C	C	C	NO
03	MIA	univs.	no	C	C	C	NO
03	MER	univs.	nivell D	C	C	C	NO
03	RAS	univs.	no	C	C	C	NO
03	OSE	univs.	fil.cat.	C	C	C	SÍ
03	TON	secunds.	no	C	C	C	NO
04	ANA	---	no	C	C	C	NO
04	COC	univs.	no	C	C	C	NO
04	JOJ	univs.	---	C	C	C	NO
04	MAT	secunds.	no	C	C	C	NO
04	MEC	univs.	fil.cat.	C	C	C	SÍ
04	RON	univs.	reciclatge	C	C	C	NO
04	ONI	secunds.	no	C	C	C	NO
05	EUU	univs.	no	C	C	C	NO
05	MJJ	univs.	no	C	C	C	NO
05	PUY	secunds.	no	CE	C	CE	NO
05	TRS	univs.	fil.cat.	CE	C	C	SÍ
06	ANI	secunds.	no	C	C	C	NO
06	BEB	univs.	no	C	C	C	NO
06	NIA	univs.	fil.cat.	C	C	C	SÍ
06	MAM	secunds.	no	C	C	C	NO
07	CME	univs.	fil.cat.	C	C	C	SÍ
07	MMM	prim.	no	C	C	C	NO
07	PPP	prim.	no	C	C	C	NO
08	FSC	univs.	fil.cat.	C	C	C	SÍ
08	PEI	prim.	no	C	C	C	NO
08	RRR	prim.	no	C	C	C	NO
09	AVV	prim.	no	C	C	C	NO

09	DDD	prim.	no	C	C	C	NO
09	ELE	secunds.	no	CE	C	C	NO
09	LIA	secunds.	no	C	C	C	NO
09	JJJ	univs.	no	C	C	C	NO
09	IDA	prim.	no	C	C	C	NO
09	MRC	---	no	C	C	C	NO
09	MSA	univs.	no	C	C	C	NO
09	NNN	univs.	fil.cat.	C	C	C	SÍ
09	OOO	univs.	nivell D	C	C	C	NO
09	SER	univs.	mòdul II	C	C	C	NO
10	ALL	univs.	nivell C	C	C	C	NO
10	GEM	secunds.	no	C	C	C	NO
10	JON	prim.	no	C	C	C	NO
10	MAG	prim.	no	C	C	C	NO
10	RIA	prim.	no	C	C	C	NO
10	SEP	prim.	no	C	C	C	NO
10	VIE	secunds.	nivell C	C	C	C	NO

Característiques socioculturals dels parlants (2)

**Mostra:** Codi de dos dígit que serveix per identificar la mostra.

**Parlant:** Codi de tres caràcters que s'utilitza per identificar el parlant.

**Edat:** Edat.

**S:** Sexe. Les respostes possibles són **d** (dona) o **h** (home).

**Població:** Població de naixement.

**Ocupació:** Ocupació.

**Grup social:** Grup social. S'ha seguit la proposta de Subirats *et al.* (1992), veg. 2.2.3.

**Estudis:** Nivell d'estudis. Les respostes possibles són **univs.** (universitaris), **secunds.** (secundaris), **prim.** (primaris), o --- (irrellevant).<sup>17</sup>

**Català:** Estudis de català. Les respostes possibles són:

— **fil.cat.:** persona que està cursant estudis de filologia catalana; la persona que té aquests estudis és la mateixa transscriptora i enregistratora de la interacció;

— **nivell C, nivell D, mòdul II, reciclatge:** persones que han assolit els nivells de coneixement de català que s'especifiquen a la taula mitjançant

<sup>17</sup> Aquesta possibilitat és comuna a altres camps de la taula.



uns cursos o exàmens determinats de l'Institut de Ciències de l'Educació, de la Junta Permanent de Català o d'altres organismes;

— **no**: persones que no tenen estudis específics de llengua catalana; en aquest grup també s'inclouen, sense especificar-ho, aquelles persones que tenen estudis primaris o secundaris que exigien un grau determinat de coneixement de català.

**LPF, LDA, LHA**: Llengua de provinença familiar, llengua de domini actual i llengua habitual actual, respectivament. Les respostes possibles són **C** (català), **E** (castellà) o **CE** (català i castellà donant preferència al català).

**CO**: Aquest camp fa referència al fet que el parlant en qüestió sap o desconeix que la interacció en la qual participa s'està enregistrant. Per tant, una resposta afirmativa (**SÍ**) significa que un individu determinat és conscient d'aquesta circumstància, i una de negativa (**NO**) vol dir que el parlant no sabia que estava sent enregistrat.

#### Annex 6. Treballs d'investigació i publicacions

Bladas Martí, Òscar (1998): *Rutines de parla. Funció i situació dins l'enunciat*. Treball d'investigació del programa de doctorat 1996-1998. Departament de Filologia Catalana. Universitat de Barcelona.

— (1999): «Les rutines de parla en el català col·loquial». Jornades de Fraseologia Comparada. Universitat Jaume I, Castelló de la Plana (en premsa).

Castellà, Josep Maria (1996): *Oralitat i escriptura en el discurs acadèmic. La complexitat lingüística en el discurs oral i escrit*. Treball de recerca per al programa de doctorat «Lèxic i discurs». Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.

Espuny Montserrat, Janina (1996): *Étude de la diaphonie dans des dialogues en face à face*. Tesi doctoral. Departament de Filologia Francesa. Universitat de Barcelona.

Oller Piqué, Anna (1996): «Uso de elementos temporales de cohesión en cuentos orales infantiles». IX Encontro da Associação de Xoves Lingüistas. A Coruña: Universidad da Coruña.

— (1998): «Una muestra de investigación en corpus orales informatizados: el *llavors* (entonces) col·loquial, cortesía en un discurso

aparentemente inofensivo». I Simposio Internacional de Análisis del Discurso. Madrid: Universidad Complutense de Madrid (en premsa).

— (1998): *El llavors col·loquial, aspectes gramaticals i pragmàtics*. Tesi de llicenciatura. Departament de Filologia Catalana. Universitat de Barcelona.

— (1999): «Connectors i fraseologia: el *llavors* col·loquial». Jornades de Fraseologia Comparada. Universitat Jaume I, Castelló de la Plana (en premsa).

Payrató, L., E. Boix, M.R. Lloret i M. Lorente (eds.) (1996): *Corpus, corpora*. Barcelona: PPU-Universitat de Barcelona.

Payrató, L., E. Boix, M.R. Lloret i M. Lorente (1996): «The *Corpus de la Universitat de Barcelona* (CUB): Getting access to spoken catalan. 5<sup>th</sup> International Pragmatics Conference, Mèxic.

Payrató, L. (ed.) (1998): *Oralment. Estudis de varació funcional*. Barcelona: Publicaciones de l'Abadia de Montserrat-Universitat de Barcelona.

Torres Vilatarsana, Marta (1999): *Les interjeccions: aproximació pragmàtica a partir de converses presencials i virtuals*. Tesi de llicenciatura. Departament de Filologia Catalana, Universitat de Barcelona.