THE ORAL CORPORA IN CATALAN AND RESEARCH INTO CASUAL CATALAN: THE CORPORA OF CONVERSATION AND FUNCTIONAL VARIETIES OF THE *CORPUS DE CATALÀ CONTEMPORANI DE LA UNIVERSITAT DE BARCELONA* (CCCUB)*

ELS CORPUS ORALS I EN CATALÀ I EL CATALÀ CORRENT: ELS CORPUS DE CONVERSA I DE VARIETATS FUNCIO‐ NALS DEL CORPUS DEL CATALÀ CONTEMPORANI DE LA UNIVERSITAT DE BARCELONA (CCCUB)

Lluís Payrató & Neus Nogué
Universitat de Barcelona
payrató@ub.edu / nnogue@ub.edu

Abstract: The aim of this study is to present a set of corpora of casual Catalan (especially oral and colloquial) gathered together in the VARCOM and PRAGMAESTIL projects carried out over the past twenty years at the University of Barcelona. The research has been conducted by the *Grup d'Estudi de la Variació* (GEV-UB), in particular by the *Grup d'Estudis de Pragmàtica i Anàlisi del Discurs* (GrEPAD). This paper reviews an oral corpus of colloquial conversation (COC, 2002), an oral corpus of registers (COR, 2004) and an audiovisual oral corpus of textual varieties (CAP, 2008). In the CAP, for the first time the images of the interlocutors are shown as they speak, alongside the transcribed texts. This paper focuses on the qualitative and research-oriented nature of these corpora. The final part of the text underlines the research possibilities that the corpora offer and also highlights the gaps remaining in the field of corpus linguistics applied to Catalan and its oral varieties.

Key words: corpus, Catalan, language varieties, oral language, applied linguistics.

Resum: Aquest treball pretén exposar el conjunt de corpus i de recerques que s'han elaborat al llarg de diferents projectes promoguts amb l'objectiu de constituir corpus representatius del català corrent (sobretot oral i col·loquial). Els projectes a què es fa referència (VARCOM i PRAGMAESTIL)

Lluís Payrató & Neus Nogué
The oral corpora in Catalan and research into casual Catalan:
the corpora of conversation and functional varieties of the CCCUB

s'han dut a terme des de final del segle xx i des de la Universitat de Barcelona, més en concret amb la feina desenvolupada per diferents subgrups del *Grup d'Estudi de la Variació* (GEV-UB), en particular pel *Grup d'Estudis de Pragmàtica i Anàlisi del Discurs* (GrEPAD). El treball revisa en especial un corpus oral de conversa col·loquial (COC, 2002), un corpus oral de registres (COR, 2004) i un corpus oral audiovisual de gèneres textuals (CAP, 2008), en què per primera vegada se suma, a la veu paral·lelitzada amb els textos transcrits, la imatge dels interlocutors. El present treball incideix en especial en el caràcter volgudament qualitatiu i orientat a la recerca d'aquests corpus, i alhora en les possibilitats d'investigació que ofereixen i en les mancances que encara són prou evidents en l'àmbit de la lingüística de corpus aplicada al català i a les seves varietats orals.

Paraules clau: corpus, català, varietats lingüístiques, llengua oral, lingüística aplicada.

❧   ❧   ❧

1. INTRODUCTION. THE VARCOM AND THE PRAGMAESTIL PROJECTS AS RESEARCH BACKGROUND

The *Corpus de Català Contemporani de la Universitat de Barcelona* (CCCUB) was created as the initial phase of the research on language variation and discourse styles developed in the VARCOM and the PRAGMAESTIL projects.

VARCOM is an abbreviation of «Variation, multimodal communication and multilingualism: discursive styles and linguistic consciousness in the production of oral texts».[1] The specific goals of VARCOM were to analyze the following aspects:

(1)   Design of discourse style typologies in multilingual speakers including verbal and non-verbal (sub)categories.

(2)   Contrastive analysis of (non-)verbal stylistic and expressive resources used in oral texts in various formats (narration, description, exposition, argumentation, instruction).

(3)   Description of synchronization and interaction mechanisms between communicative modalities (verbal, vocal and non-verbal).

Lluís Payrató & Neus Nogué
The oral corpora in Catalan and research into casual Catalan:
the corpora of conversation and functional varieties of the CCCUB

(4)   Study of language awareness in multilingual speakers and their expectations about multilingualism in personal and professional domains.

(5)   Analysis of factors that determine the variation in discourse styles in the oral production of multilingual speakers (Catalan, Spanish, and English).

PRAGMAESTIL is an abbreviation of «Pragmatics, style and identities: analysis of verbal and non-verbal properties in multilingual speaker discourse».[2] Understanding style as a manifestation of the individual identity that becomes concrete in the contextualized production of texts, the main goal of the PRAGMAESTIL project was to analyze some of the most significant aspects in the intralinguistic (in Catalan) and interlinguistic (in Catalan, Spanish, English/German) construction of discursive styles by a selected group of multilingual speakers. Informants were, on the one hand, bilingual Catalan-Spanish speakers with one of the two languages as their L1 and family language and, on the other, speakers raised in bilingual homes, all of whom have knowledge of English as L3.

From the dual perspective described above, the PRAGMAESTIL project specifically aims to determine the style types constructed in relation to a standard typology (particularly according to standard text types: narrative, descriptive, expository, argumentative and directive) and to establish the types constructed from categories proposed during the analysis. The specific goals of PRAGMAESTIL were to analyze the following aspects:

(1)   Logico-semantic or argumentative connectors and pragmatic connectors.

(2)   Thematic textual progression in relation to text structure and type (narrative, descriptive, expository, argumentative and directive).

(3)   Verbal and non-verbal expression of deictic elements (person, space, time, and manner) and of the categories related to facts/actions, statements/questions and spatial references.

Lluís Payrató & Neus Nogué
The oral corpora in Catalan and research into casual Catalan:
the corpora of conversation and functional varieties of the CCCUB

(4)   Idiolectal markers of speaker identity (lexical selection, construction and syntactic complexity, cohesion formulas, degree of assertiveness, involvement, formality and politeness, among other factors).

2. THE *CORPUS DE CATALÀ CONTEMPORANI DE LA UNIVERSITAT DE BARCELONA* (CCCUB)

The «corpus of contemporary Catalan of the University of Barcelona» (CC-CUB) is the result of several projects added to VARCOM and PRAGMAESTIL. Its aim is to collect, transcribe, publish and analyse a corpus of contemporary spoken and written Catalan.[3]

The CCCUB covers a number of geographical, sociocultural, and functional varieties of current Catalan. It has been designed as a set of interconnected modules, each one focusing on a particular dimension of linguistic variation. While some of these modules have already been completed, and already are available for research, others are still in process.

The module covering geographical variation (Corpus Oral Dialectal, COD) has compiled morphological and textual data from fieldwork carried out at 86 locations throughout the Catalan speaking area. The data include 403,766 morphophonological forms, and 543 recorded samples of casual speech, each lasting 10 minutes (vid. <http://www.ub.edu/cccub/corpusoraldialectal-cod.html>).

The sociocultural module (Corpus Oral Social, COS, vid. Boix-Fuster *et al.* 2006) is designed to contain 78 face-to-face interviews, and constitutes a sample of the social varieties currently spoken in the area of Barcelona. Some of the informants are Catalan speakers married to Spanish speakers, which gives additional evidence for language contact studies.

An overview of functional variation is obtained by means of a corpus of registers (Corpus Oral de Registres, COR). Moreover, some specific functional varieties have received particular attention: casual speech (Corpus Oral de Conversa Col·loquial, COC), oral publicity (Corpus Oral de Publicitat, COP), radio newscasting (Corpus

---

3. Members of the original CCCUB Project were Núria Alturo, Òscar Bladas, Emili Boix, Esteve Clua, Mireia Galindo, M. Rosa Lloret, Mar Massanell, Anna Oller, Marta Payà, Lluís Payrató, M. Pilar Perea, Clàudia Pons, Marta Torres, Joaquim Viaplana, F. Xavier Vila (Universitat de Barcelona) i Amadeu Viana (Universitat de Lleida).

LLUÍS PAYRATÓ & NEUS NOGUÉ
The oral corpora in Catalan and research into casual Catalan:
the corpora of conversation and functional varieties of the CCCUB

d'Informatius Orals, CIO), and written media discourse (Corpus Escrit del Català Actual, CECA).

The whole corpus and the organization of these modules can be seen in Figure 1.
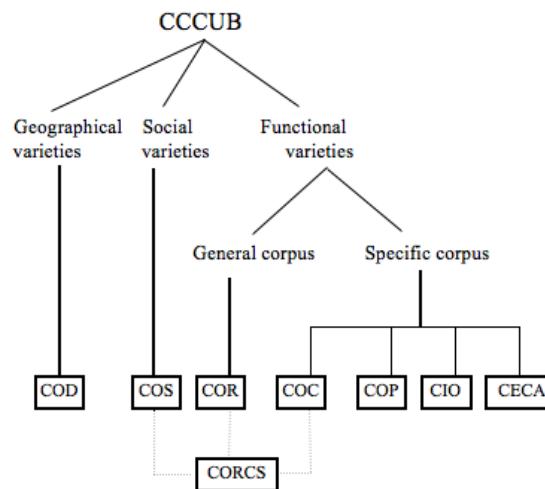
Figure 1. Modules of the *Corpus de Català Contemporani de la Universitat de Barcelona* (CCCUB)

3. THE *CORPUS ORAL DE CONVERSA* (COC)

The basic goal of the «oral corpus of (colloquial) conversation» (COC) was to produce suitable materials for the analysis of Catalan conversation and colloquial register (spontaneous, interactive and informal spoken language). It contains 10 conversations, each of which lasts 30 minutes (a total of 5 hours/70,000 words). It is available as a CD-ROM, with an accompanying book (Payrató & Alturo 2002) that contains a complete discourse transcription and partial phonetic transcription, and also on the Internet (<http://www.ub.edu/cccub/corpusoraldeconversacolloquial-coc. html>), where different files with the recording and the transcription are available. The transcription is synchronized with the voice recording. The corpus also contains ethnographic information about the communicative situation in which each conversation takes place and tables with details about the sociocultural characteristics of the interlocutors. In addition to the edited texts, 40 conversations of 30 minutes each are also available in the Department as complementary data.

Lluís Payrató & Neus Nogué
*The oral corpora in Catalan and research into casual Catalan:
the corpora of conversation and functional varieties of the CCCUB*

3.1 TRANSCRIPTION CONVENTIONS AND TREATMENT OF THE DATA

The data are recorded in a discourse transcription with the addition of some aspects of intonation and other features of non-verbal vocal behaviour, according to the conventions described in this subsection (see also Du Bois *et al.* 1993; Payrató 1995 Payrató & Alturo 2002; Bladas 2009). For this corpus and subsequent ones (COR and CAP), an important methodological decision regarding transcription was taken: instead of developing a new transcription system, which would contribute to an even greater fragmentation of a weak international standard, the system of Santa Barbara Corpus of Spoken American English was adopted, though very slightly modified (cf. Bladas 2009; Du Bois *et al.* 1993, and <https://www.linguistics.ucsb.edu/research/santa-barbara-corpus>).

3.1.1 Prosodic aspects

Final tone sequence	
descending final tone	\
ascending final tone	/
maintaining final tone	–
Truncated tonal group	--
Maintenance of the intonation	
ascending	{(A) affected text}
descending	{(B) affected text}
Altering the voice	{(EV) affected text}
Intensity	
high	{(F) affected text}
very high	{(FF) affected text}
low	{(P) affected text}
very low	{(PP) affected text}
Time	
fast	{(AC) affected text}
slow	{(DC) affected text}
Short, medium and long lengthening	: :: :::

Lluís Payrató & Neus Nogué
The oral corpora in Catalan and research into casual Catalan:
the corpora of conversation and functional varieties of the CCCUB

3.1.2 Vocal aspects

Simultaneous laughter with speech	{(@) affected text}
Non-simultaneous laughter with speech	
one symbol per syllable	@
long fragment with duration	@R(duration)R@

Inhalation and exhalation   (INH)(EXH)

3.1.3 Pauses and overlaps

Pause
very short ($0.1 < p < 0.3$)	(. duration)
short ($p < 1$)	(.. duration)
medium ($1 \leq p < 3$)	(… duration)
long ($p > 3$)	(…. duration)

Overlapping   [affected text]

3.1.4 Regularizations and comments

Elision   (spelling of the elided sound)

Transcriber's comment
descriptive	(comment)
general	((comment))

3.1.5 Problematic fragments

Doubt about the transcribed text   {(??) affected text}

Unintelligible fragment
one sign per syllable	x
long fragment with duration	xX(duration)Xx

Lluís Payrató & Neus Nogué
The oral corpora in Catalan and research into casual Catalan:
the corpora of conversation and functional varieties of the CCCUB

3.1.6 Other aspects

Second language   {(L2) affected text}

Truncation of word   -

Truncation of word and of tonal group   - --

3.2. MAIN FEATURES OF THE SAMPLE AND THE SPEAKERS[4]

The COC contains ten recordings of casual or colloquial conversation (interactive tenor, informal tone, spontaneous oral channel and general theme) in a friendly or family situation made in the years 1992, 1993, 1994 and 1997, as reflected in Table 1.

Table 1. Sample, location, and date of the recording

Mostra	Situació	Data
01	dinner in the dining room	1993
02	chatting on the couch after a meal out	14/2/93
03	lunch in the dining room	DG/5/93
04	lunch	19/2/93
05	chatting on the couch after a football match	nit/8/1997
06	dinner and coffee inthe dinning room	1992-93
07	before dinner in the dining room	10/11/93
08	after-dinner conversation on the sofa with the television on	1994
09	after-dinner conversation in the dinning room	8/12/93
10	courtesy visit in the dining room	8/12/93

This corpus comprises approximately 281 minutes (28 minutes on average per text or recording) and 70493 words (7049 words on average), counting on the initials of the speakers and all the symbols established in the transcription system.

4. This subsection reproduces data included in Payrató & Alturo (2002).

Lluís Payrató & Neus Nogué
*The oral corpora in Catalan and research into casual Catalan:
the corpora of conversation and functional varieties of the CCCUB*

Each recording lasts between 23 and 31 minutes, and contains between 5200 and 8966 words, as can be seen in Table 2.

Table 2. Specification of the length and the number of words of the COC

Sample	Minutes	Words
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
TOTAL	281,67	70493

Due to the high number of participants, the technical means used (a conventional recorder) and the surreptitious nature of the recording, the sound quality of the samples is not particularly good. Informants were adult speakers of central dialects (especially in the region of Barcelona) who, in general, do not have a specialist knowledge of Catalan.

Regarding the geographical distribution of speakers, there were 23 people from Barcelona (Barcelonès, 40.4%) and 34 from elsewhere (59.6%). Among those not from Barcelona, eight were from Mataró (Maresme), six from Badalona (Barcelonès), five from Esparreguera (Baix Llobregat), three from Vic (Osona), three from Caldes de Montbui (Vallès Occidental), two from Sant Martí Sescorts (Osona), 1 from Susqueda (La Selva), 1 from Vespella (Osona), 1 from Sant Feliu de Codines (Vallès Oriental), 1 from Rupià (Baix Empordà), 1 from Capellades (Anoia), 1 from Premià (Maresme) and 1 from Manlleu (Osona). This distribution is shown in Figure 2.

Lluís Payrató & Neus Nogué
The oral corpora in Catalan and research into casual Catalan:
*the corpora of conversation and functional varieties of the CCCUB*
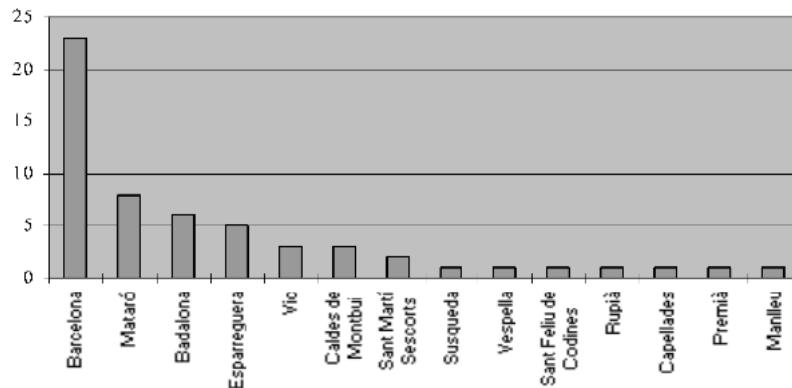
Figure 2. Geographical distribution of COC speakers by populations

By counties, 29 people were from the Barcelonès (50.9%), nine from the Maresme (15.8%), seven from Osona (12.3%), five from Baix Llobregat (8, 8%), three from the Vallès Occidental (5.3%), one from the Vallès Oriental (1.7%), one from the Baix Empordà (1.7%), one from La Selva (1.7%) and one from Anoia (1.7%). Therefore, in this part of the COC there is a clear predominance of the speakers from the Barcelonès, especially around Barcelona: the Maresme, the Baix Llobregat, the Vallès Occidental and the Vallès Oriental (47 speakers, or 82.45% of the total). The remaining ten (17.54%) were from outside this region, specifically from Osona, Baix Empordà, La Selva and Anoia.

Regarding the sex of the speakers, 35 out of the 57 were women (61.4%), and 22 (38.5%) were men. Regarding the social stratification of speakers, two major groups are established:[5]

a) 30 middle-class speakers (52.63%), 28 of whom were employees or middle range technical staff (49.12%), and two senior or professional technical staff (3.5%).
b) 25 working class speakers (43.85%), 18 of whom were manual workers (31.57%), six old manual workers (10.52%), and one a supervisor (1.75%).

No social group was stipulated in the case of two children (3.5%) due to lack of data and because their linguistic or communicative participation in the conversations was minimal. The date are presented in chart form in Figure 3.

5. This information is based on the proposal of the Survey of the Metropolitan Region of Barcelona (Subirats *et al.* 1992). For more detailed information about the geographical and social characteristics of the speakers, see Annex 2 of Payrató & Alturo (2002: 20-21).

Lluís Payrató & Neus Nogué

*The oral corpora in Catalan and research into casual Catalan:
the corpora of conversation and functional varieties of the CCCUB*
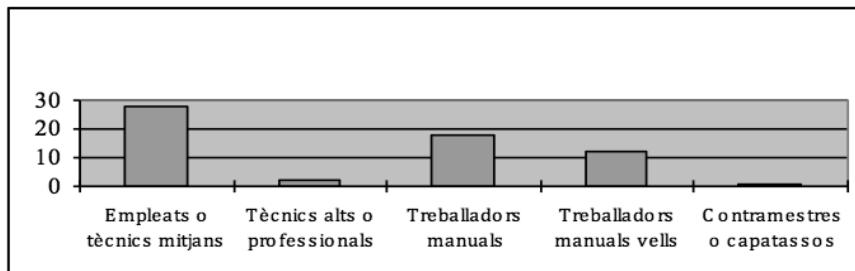
Figure 3. Social distribution of the speakers of the COC

3.3 SAMPLE

The on-screen presentation of the CD is shown in Image 1, which reproduces the beginning of conversation 1.

1 DIT quina po::r\	20 MMA (.. 0.61) (INH)
2 ((en les línies següents se senten de fons diversos	21 MMA i el:--
sorolls d'olles, coberts i plats))	22 MMA la mama què ha dit/
3 MMA (.. 0.83) pensar que m'anava a cremar\	23 MM? (.. 0.78) (INH)
4 DIT (.. 0.69) sí/	24 REP i no hi [entenc] jo nen\
5 REP i el--	25 DI? [xx]
6 REP {(@) i el vicari de_	26 REP [no hi en]tenc\
7 REP de Badalona que es pixa i no se n'adona\}	27 MM? [{(P) x}]
8 MMA [1 el què:/]	28 MMA (.. 0.87) {(P) ha\
9 REP [1 @@@@][2 @@]	29 MMA ha ha ha ha ha\}
10 DIT [2 @@]@@@@ (INH)	30 REP {(P) ai senyor\}
11 LAU (.. 0.30) i {(??) [1 jo estic a]quí--}	31 LAU (... 2.81) ai senyor\}
12 MMA [1 què:/]	32 MMA (.. 0.52) (INH)
13 MMA xx]	33 MMA no\
14 DIT @@@@@@@@[2 @@]@@ (INH)	34 MMA el vicari d'Olot que té pèls al ninot\
15 MMA [2 què:/]	35 (... 1.14)
16 DIT {(@) [3 el] vicari de Barcelona que es pixa i no	36 REP {(??) i se sentia] [si és arregla]da\
se n'adona\}	37 DIT [@@@@]
17 REP [3 x]	38 (... 1.81)
18 DI? (.. 0.40) {(P) @@}	39 REP be[neïm\]
19 REP (.. 0.90) que es pixa i no se n'adona\	40 ??? [x]
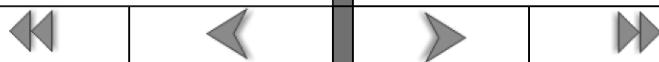
Image 1. Reproduction of the screen access to the COC corpus (conversation 1)

Lluís Payrató & Neus Nogué
The oral corpora in Catalan and research into casual Catalan:
the corpora of conversation and functional varieties of the CCCUB

From this screen, the soundtrack (synchronized with the text), the phonetic transcription of the text, the ethnographic information on the communicative situation and the basic characteristics of speakers (age, origin, profession, social group) can be accessed by clicking on buttons and boxes (not shown in the image). In the printed book (Payrató & Alturo 2002), the text is shown in parallel columns (with discursive and phonetic transcription) of the first five minutes of each conversation (see table 3).

Table 3. Presentation of the corpus (example of a fragment of conversation 7)

Lluís Payrató & Neus Nogué
*The oral corpora in Catalan and research into casual Catalan:
the corpora of conversation and functional varieties of the CCCUB*

4. THE CORPUS ORAL DE REGISTRES (COR)

The basic goal of the «oral corpus of registers» (COR) was to produce suitable materials for the comparative analysis of Catalan registers and typical speech events in the studied community. It includes 26 speech events and 168 texts of different length (a total of 9 hours/154,000 words, approximately) and is also available as a CD-ROM, with an accompanying book (Alturo *et al.* 2004) containing complete discourse transcriptions, and also on the Internet (<http://www.ub.edu/cccub/corpusoralderegistrescor.html>). As above, the transcribed text in the CD-ROM and the voice recording are synchronized. The corpus is structured into nine areas of use or domains (judicial and civil, business, education, politics, culture, religion, media, private communication and provoked interactions) and 26 genres (presentation, sermon, story, recipe, etc.). Ethnographic information about the communicative situation in which each speech event takes place is included, as well as tables containing details about the sociocultural characteristics of the interlocutors. The transcription applies the same conventions as in the COC (see 3.1), but this time without phonetic transcription.

4.1 ITEMS AND STRUCTURE

The COR is structured in fields and in discourse genres.[6] By *domain* we understand the sociocultural context in which a particular discourse is produced, such as education, business, and so on. The term *discourse* or *speech genre* refers to the communicative event occurring in a particular social context, for example a press conference or the message left on an answering machine.

The publication comprises a total of 168 texts that make up a sample of 26 speech genres in seven different fields (see table 4). This sample aims to offer a diverse, heterogeneous body of reference material for the study of functional variation in contemporary oral Catalan. The CD-ROM contains the full texts, which can be consulted directly or through any word processing program. The printed edition includes only brief fragments of each genre. The set of materials and the structure of the corpus are reproduced in Table 4.

6. This subsection reproduces data included in Alturo *et al.* (2004: 5).

Lluís Payrató & Neus Nogué
The oral corpora in Catalan and research into casual Catalan:
the corpora of conversation and functional varieties of the CCCUB

Table 4. Items and structure of the COR

CODE	DOMAIN	DISCOURSE GENRE	TEXTS	WORDS	MINUTES
CIV1	CIVIL	Courtroom trial	1	950	4,52
CIV2		Civil wedding	1	3097	21,51
EDUC1	EDUCATIO-NAL-ACADEMIC-SCIENTIFIC	Teachers meeting	1	9579	33,45
EDUC2		School story	1	2211	7,55
EDUC3		Master class	1	7978	30,25
EDUC4		Conference	1	1393	25,50
POL1	POLITIC	Political meeting	1	6268	30
POL2		Parliamentary session	1	6587	32
POL3		Plenary session of a town hall	1	4413	40,38
POL4		Press conference	1	13643	51
PRIV1	PRIVATE	Familiar story	1	1140	4,22
PRIV2		Messages on an answering machine	100	7066	40,45
PRIV3		Telephone conversation	1	11081	35,50
EMP1	BUSINESS-COMMER-CIAL	Interview of practices	1	4560	22
EMP2		Technical talk	1	4968	30,15
EMP3		Fish selling	1	7202	33,30
EMP4		Hiring a trip	1	4940	17,35
EMP5		Market sale	25	7267	26,56
EMP6		Telephone inquiries	20	8972	30,20
CULT1	CULTURAL	Presentation of a video	1	4278	27,40
CULT2		Guided tour	1	8943	30
CULT3		Theater company rehearsal	1	9351	30
CULT4		Poetry recital	1	4723	46
REL1	RELIGIOUS	Sunday sermon	1	2485	11,15
REL2		Wedding	1	7011	34
REL3		Mass	1	4092	27
			TOTAL = 168	**TOTAL = 154198**	**TOTAL = 721,44**

Lluís Payrató & Neus Nogué
The oral corpora in Catalan and research into casual Catalan:
the corpora of conversation and functional varieties of the CCCUB

4.2 samples

We now provide three samples corresponding to three areas: political (and public), cultural and private. The first is the beginning of a plenary session of a city council, the second a play rehearsal by a non-professional theatre group, and the third the telling of a story in a school setting.

For the purposes of contextualization, data regarding each extract are provided, including information about the interlocutors (the personal or professional relationship between them, the role they play in the interaction, age, social class, etc.), the setting, and any other relevant features in each communicative exchange.

4.2.1 Plenary session of a town hall

Plenary session of the Town Hall of Arenys de Munt held on March 4, 1997. The session is held in a meeting room of the town hall, where there is a round table and shelves with books. There are seven participants, six of home are men between 30 and 50 years old (SOS, mayor and president of the plenary; CAA, town councilor; REE, councilor for culture; SOO, councilor for government; DRR, municipal group spokesman, and HHG, councilor), and one is a 28-year-old woman (NTT, secretary of the plenary session). The event is open to the public and broadcast live on local radio.

The plenary president opens the session and puts the minutes of the previous session to a vote. All the councilors have previously received the minutes in writing. Next, the president presents the points of the agenda that must be dealt with; then, after questions (submitted previously in writing to the mayor), he puts the proposals to a vote.

1.	PDR:	tothom_
2.		present_
3.		està:_
4.		a punt de començar aquest ple_
5.		comença el ple\
6.	SOS:	(.. 0.56) començaríem_
7.		doncs_
8.		la sessió::_
9.		(. 0.24) del ple_
10.		(.. 0.60) {(AC) corresponent a avui dimarts_
11.		dia_}
12.		e:_

Lluís Payrató & Neus Nogué

The oral corpora in Catalan and research into casual Catalan:
the corpora of conversation and functional varieties of the CCCUB

13.		(. 0.28) {(AC) dia quatre de març de mil nou-cents noranta-set\
14.		(.. 0.84) el primer punt de l'ordre del dia_
15.		és l'aprovació_
16.		si s'escau_
17.		de l'esborrany_
18.		(.. 0.88) de l'acta de la sessió anterior\}
19.		(… 1.52) en aquí tothom_
20.		x--
21.		ha rebut l'acta_
22.		so- --
23.		si hi han algunes esmenes a fer_
24.		(… 1.06) sembla que:_
25.		(… 1.67) s'aprova\
26.	???:	{(P) xxx\}
27.	SOS:	(.. 0.83) {(AC) el segon punt_
28.		és l'aprovació de l'expedie:nt_
29.		de contribucions especials_
30.		(.. 0.73) la_
31.		de {(??) l'urbanització} de l'avinguda de la Pau\ (AC)}
32.		(. 0.27) un tros de:_
33.		(.. 0.86) de carrer que va des de la_
34.		de la plaça de l'avinguda de la Pau_
35.		fins al_
36.		(.. 0.67) al bloc del {(L2) Barcino\}
37.	CAA:	(… 1.07) (estossec)
38.		bé\
39.		tenim aprovat ja_
40.		com sabe- --
41.		com saben_
42.		el projecte:_
43.		e::_
44.		per al ple_
45.		(.. 0.54) de l'urbanització_
46.		del sector_
47.		de l'avinguda de la Pau\
48.		(. 0.24) i avui es posa_
49.		a aprovació del ple_
50.		(.. 0.34) l'expedient_
51.		de contribucions especials_
52.		de:_
53.		per paga:r_

Lluís Payrató & Neus Nogué
The oral corpora in Catalan and research into casual Catalan:
the corpora of conversation and functional varieties of the CCCUB

54.   perquè {(F) puguin} pagar els veïns afectats\
55.   (.. 0.85) la quota que es fixa_
56.   (.. 0.55) és_
57.   un sixanta per cent_
58.   els veïns_
59.   i un coranta per cent_
60.   a pagar l'ajuntament\
61.   (… 1.68) ha passat_
62.   per la comissió informativa_
63.   corresponent_
64.   (. 0.12) i avui_
65.   e::_
66.   és lo que posem_
67.   a aprovació del ple\
68.   (.. 0.42) no és_
69.   (.. 0.45) no és una cosa complicada_
70.   és una cosa_
71.   doncs_
72.   e::_
73.   habitualment_
74.   doncs es fa_
75.   per contribucions especials_
76.   (INH)
77.   (.. 0.31) tot lo que són urbanitzacions_
78.   perquè {(F) milloren} les finques_
79.   (.. 0.48) i:_
80.   creiem_
81.   que aquesta és una bona_
82.   mesura_
83.   que hem anat fent_
84.   ane la::_
85.   a Arenys de Munt_
86.   com hem anat urbanitzant carrers_
87.   (.. 0.48) i_
88.   (. 0.25) m:_
89.   ens ha dat un bon_
90.   un bon_
91.   e:_
92.   servei_
93.   i un bon_
94.   (. 0.28) i un bon fer\

Lluís Payrató & Neus Nogué
The oral corpora in Catalan and research into casual Catalan:
the corpora of conversation and functional varieties of the CCCUB

95.   (… 1.30) fins avui\
96.   (… 2.26) això és lo que es posa_
97.   a:_
98.   a consideració del ple_
99.   perquè s'aprovin_
100.   doncs el_
101.   la:_
102.   (. 0.25) {(F) imposició_}
103.   (. 0.26) d'aquestes contribucions especials\
104.   (… 2.05) {(P) si hi ha alguna:_}
105. DRR: sí\
106.   nosaltres_
107. ???:   x--
108. DRR: d'entrada_
109.   volíem fer_
110.   una pregunta_
111.   i és si_
112.   els veïns_
113.   afectats_
114.   per aquestes contribucions_
115.   (INH)
116.   han estat reunits_
117.   i si estan d'acord amb_
118.   (. 0.25) am(b) aquesta distribució:_

4.2.2 Political meeting

Meeting of Esquerra Republicana de Catalunya at the Palau de Congressos in Barcelona on February 29, 1996, in the electoral campaign before the elections of March 3 of the same year. Eleven speakers (candidates, members of the executive council and a guest from Eusko Alkartasuna), as well as the director of the event, take part. Thirty minutes of the meeting are transcribed, corresponding to interventions by Albert Roig (ARR, Vilanova i la Geltrú, 31 years old, Catalan-speaking family, primary school teacher and regional president of ERC del Penedès-Anoia), Begoña Lasagabaster Olazábal (BLL, Irun, 34 years old, Basque-speaking family, a graduated in European studies at the College of Europe, Legal Department, Bruges, Belgium, Congressman by Eusko Alkartasuna), and Pilar Rahola Martínez (PPR, Barcelona, 38, Catalan-speaking family, graduated in Hispanic and Catalan philology, fourth Deputy Mayor of the City Council of Barcelona, responsible for commerce, delegated to the

Lluís Payrató & Neus Nogué
*The oral corpora in Catalan and research into casual Catalan:
the corpora of conversation and functional varieties of the CCCUB*

Congress for Esquerra Republicana de Catalunya). The presentations made were also transcribed. The director of the event is Josep Huguet Biosca (JHH, Manresa, 45 years old, Catalan-speaking family, degrees in engineering and a history degree, member of parliament).

1.		((aplaudiments))
2.	JHH:	en els debats s'han parla:t_
3.		i parla:t_
4.		i parla:t_
5.		de la crisi de l'estat del benestar\
6.		(. 0.22) però {(F) només} Esquerra Republicana_
7.		ha posat sobre la taula_
8.		(.. 0.45) el fet de que:_
9.		el nostre esfo:rç_
10.		am(b) el nostre esforç fiscal_
11.		(.. 0.45) de fet_
12.		podríem mantenir_
13.		i millorar_
14.		(. 0.18) el nostre estat del benestar\
15.		(.. 0.57) {(F) qui} sinó Esquerra Republicana defensarà a Madrid_
16.		(.. 0.61) l'estat català del benestar\
17.		(.. 0.38) oi Albert\
18.		(.. 0.43) amb vosaltres_
19.		Albert Roig\
20.		(. 0.16) el segon candidat per la llista de Barcelona_
21.		{(AC)(F) i que amb la vostra ajuda_ ((aplaudiments del públic))
22.		hem de intentar portar també a Madrid\}
23.		((11 segons: mentre el públic aplaudeix, ARR s'aixeca i puja a l'escenari))
24.	ARR:	bona nit catalans_
25.		(. 0.20) i catalanes\
26.		(.. 0.48) de debò\
27.		(.. 0.66) de debò\
28.		(.. 0.51) avui ens han respos(t)\
29.		(.. 0.72) han respo:s(t)_
30.		(.. 0.59) a la nostra candidata\
31.		(… 1.88) l'inefable Trias de Bes_
32.		(.. 0.64) ha dit_
33.		(… 1.73) que:_
34.		(. 0.10) si volem la independència_
35.		(… 1.44) que molt bé\
36.		(.. 0.51) però que haurà de passar per Madrid\

Lluís Payrató & Neus Nogué
The oral corpora in Catalan and research into casual Catalan:
the corpora of conversation and functional varieties of the CCCUB

37.   (.. 0.53) {(F) no n'han de fer re:: a Madrid del que decidim nosaltres\}
38.   (.. 0.80) nosaltres_
39.   (.. 0.33) decidim el nostre futur_
40.   (.. 0.35) i ells_
41.   (. 0.24) no n'han de fer absolutament re\
42.   (.. 0.77) i si són tan demòcrates_
43.   (.. 0.72) com sembla_
44.   (. 0.20) que volen ser_
45.   (.. 0.76) la primera mesura que prenguin\
46.   (.. 0.68) sigui\
47.   (.. 0.39) reformar la Constitució_
48.   (. 0.20) com hem anat dient durant {(F) tots} aquests dies\
49.   (… 1.00) i és més\
50.   (… 1.59) ens parlen_
51.   (… 1.08) d'una administració única\
52.   (.. 0.85) i els hi diem_
53.   (.. 0.60) un no::u_
54.   (. 0.11) {(L2) café para todos/}
55.   (. 0.11) diuen no::\
56.   (. 0.18) és que fins i tot_
57.   ens volen fotre el cafè: aquests del PP\
58.   (.. 0.34) {(F) fins i tot el cafè:\} ((aplaudiments del públic))
59.   (… 5.84) ((ARR espera que el públic pari d'aplaudir per continuar el discurs))
60.   ARR: i els altres_
61.   els que diuen que són de casa\
62.   (.. 0.38) ho deia molt bé la Marta\
63.   (.. 0.80) diuen que votar Esquerra\
64.   (.. 0.79) és llençar el vot_
65.   (.. 0.75) a la paperera\
66.   (… 2.16) que ho senti ben fort\
67.   (.. 0.55) el Molins\
68.   (… 1.28) diguem-li_
69.   (. 0.23) ben clar\
70.   (… 1.42) el vot claudicant\
71.   (… 1.32) el vo:t_
72.   (. 0.14) que es ven per un mísere_
73.   (. 0.29) quinze per cent\
74.   i ara un trenta per cent\
75.   (.. 0.33) és el vot de Convergència i Unió\
76.   (.. 0.61) el vot del futur\
77.   (.. 0.59) el vot_

Lluís Payrató & Neus Nogué
The oral corpora in Catalan and research into casual Catalan:
the corpora of conversation and functional varieties of the CCCUB

78.　　　　　que té un objectiu clar i que sap què vol per Catalunya_
79.　　　　　(.. 0.39) és el vot d'Esquerra Republicana\
80.　　　　　(. 0.27) és el vot de {(F) tots} nosaltres\
81.　　　　　(.. 0.93) {(F) nosaltres som el futur\ ((aplaudiments del públic))
82.　　　　　(.. 0.47) ells_
83.　　　　　(.. 0.47) són el passat\
84.　　　　　(.. 0.57) són la claudicació\

4.2.3 School story

Telling of the story «The puppets» inside the daily activities of a kindergarten class (3 years). The recording took place on one day in the academic year 1994-1995, in a classroom in a municipal nursery school in Barcelona. The teacher who tells the story (MOL) has a book with illustrations, which she shows to the children listening to her. Some children (NEN) interrupt the teacher to participate in the story-telling, especially one of them, who intervenes more often than the others. The teacher is a middle-class woman, 55 years old, from a Catalan-speaking Catalan family, but without any specific studies of Catalan.

1.　　NEN:　s'ha acabat\
2.　　MOL:　s'ha acabat\
3.　　NEN:　{(??) ara més\}
4.　　NEN:　xxx
5.　　NEN:　les cabretes\
6.　　MOL:　(… 1.27) una vegada:_
7.　　　　　(… 1.14) hi (ha)via una mare cabreta_
8.　　　　　(.. 0.41) que tenia_
9.　　　　　set cabridets petitons\
10.　　　　(… 1.04) vivien am una caseta al mig del bosc\
11.　　　　(.. 0.80) la mare un dia els hi va dir [an els cabri]dets_
12.　　NEN:　[el llop\]
13.　　　　　el llop\
14.　　MOL:　{(E) he d'anar a comprar perquè_
15.　　　　(.. 0.58) no tenim_
16.　　　　mo:lt poc menjar\
17.　　　　he d'anar a comprar_
18.　　　　(.. 0.66) p(e)rò vosaltres com que sou petitons_
19.　　　　jo torno de seguida_
20.　　　　(.. 0.37) no\
21.　　　　m:_
22.　　　　quedeu-se a casa_

Lluís Payrató & Neus Nogué

The oral corpora in Catalan and research into casual Catalan:
the corpora of conversation and functional varieties of the CCCUB

23.		(.. 0.49) i no obriu la porta a ningú\
24.		(.. 0.46) perquè:_
25.		(… 1.12) per aquí a f- --
26.		al--
27.		al:_
28.		bo:sc_
29.		(… 1.12) l'atre dia vaig veure córrer un llop\
30.		(. 0.24) que no us enredi\
31.		no li obriu pas la porta eh si [truca el llop\]
32.	NEN:	[{(F) el llop\}]
33.	MOL:	(.. 0.56) i les cabretes li van dir a la mare_
34.		(… 1.19) {(E) mare\
35.		no ho farem no\
36.		vés\
37.		vés\
38.		vés que nosatres jugarem molt_
39.		(. 0.26) i no obrirem la porta a ningú\
40.		encara que truquin no obri[rem a ningú\}]
41.	NEN:	[el--
42.		el ll]o::p\
43.	MOL:	{(E) doncs bé\
44.		tanca_
45.		i ara torno eh\
46.		en seguida torno a ser a casa\}
47.		i les cabretes es van posar a juga:r_
48.		a dintr- --
49.		a: caseta_
50.		molt contentes\
51.		(.. 0.77) estaven jugant molt bé molt bé_
52.	NEN:	[el llo:p\]
53.	MOL:	[amb totes les seves] joguines_
54.	NEN:	el llo[:p\]
55.	MOL:	[que] senten_
56.		(MOL dóna quatre cops, que continuaran intermitentment durant 6 línies)
57.	MOL:	(.. 0.70) {(E) ui\
58.		(.. 0.50) {(P) qui truca\}
59.		la mare ens ha dit que no obrim\ (E)}
60.		i pregunten_
61.		(.. 0.36) {(E) qui hi ha:\}
62.		(.. 0.60) i senten una {(E) veu que fa:_
63.		(.. 0.78) obriu obriu\
64.		que sóc la vostra mare\} ((se senten dos cops))
65.		{(E)(P) oh\

Lluís Payrató & Neus Nogué
The oral corpora in Catalan and research into casual Catalan:
the corpora of conversation and functional varieties of the CCCUB

66.	(.. 0.85) ai no\
67.	la mare no pot ser\
68.	(.. 0.51) que la mare no té aquesta veu\}
69.	(… 1.64) deu ser el llop\
70.	que ens ho va dir la mare\
71.	no obrirem eh\
72.	no obrirem\
73.	que la mare ens va dir que no obríssim\}
74.	(.. 0.88) i les cabretes li diuen_
75.	(.. 0.63) {(E) tu no ets la nostra mare\
76.	tu ets el llop\
77.	i no t'obrirem la porta\
78.	vés-te'n\
79.	vés-te'n\}
80.	(.. 0.48) i el llop_ ((se senten dos cops))
81.	{(E) a::h\
82.	(.. 0.69) m'heu descobert\
83.	doncs ja tornaré\
84.	ja tornaré\}
85.	(.. 0.36) i se'n va anar tot enfadat\

5. THE CORPUS AUDIOVISUAL PLURILINGÜE (CAP)

The «plurilingual audiovisual corpus» (CAP) contains audiovisual recordings and the accompanying transcriptions of 12 participants. The corpus is multilingual, with samples in Catalan, Spanish and English; multitextual, since it includes five types of text (narration, description, argumentation, exposition and instruction) and two procedures for eliciting the information (experiential and experimental); and multi-modal (it contains texts, voice recordings and images). The corpus was begun as part of the project VARCOM and was developed later under the project PRAGMAESTIL (see section 1.1).

This corpus was compiled by the research group GrEPAD (Grup d'Estudis de Pragmàtica i Anàlisi del Discurs, (see <http://www.ub.edu/grepad>), which is inte-grated into the GEV (Grup d'Estudi de la Variació, see <http://www.ub.edu/GEV>), the main research group in the Section of Catalan Language of the University of Barcelona. As can be seen in Figure 2 (see section 2), this corpus completes the series of applied materials made of the corpora COC (colloquial conversation, Payrató & Alturo 2002), COR (registers, Alturo *et al.* 2004), and COS (social varieties, Boix *et al.* 2006), all of which are part of the CCCUB.

Lluís Payrató & Neus Nogué
The oral corpora in Catalan and research into casual Catalan:
the corpora of conversation and functional varieties of the CCCUB

5.1 CONSTITUTION OF THE CORPUS. RESEARCH PROTOCOL, METHODOLOGY AND ACCESS

5.1.1 Informants

The informants were 18 women aged between 18 and 30 years who were educated in the 1980s and 1990s (with a posterior selection of 12 final informants). All of them are second-cycle students of Economics or similar subjects who were born and currently reside in the metropolitan region of Barcelona and, by extension, in the Central Catalan linguistic area, with a level of English of Cambridge First Certificate (FCE) or equivalent. According to their family tongue, they belong to one of the following groups: Catalan-speaking (with Catalan as L1), Spanish-speaking (with Spanish as L1) or raised in a bilingual home (with both Catalan and Spanish as family tongues).

5.1.2 Sessions, recordings and transcriptions

Three recording sessions took place (L1/L2/English), with a minimum of 30 days between each session and the next. Digital recording was performed using two cameras (general view/facial view). We then selected 12 of the 18 recordings, depending on factors such as fluency in the three languages, competence in English, and lack of non-verbal inhibition. The transcription follows the same conventions as COC (see 3.1), but without phonetic transcription.

5.1.3 Elicitation system

The elicitation system consisted of three semidirected interviews in which informants were requested to produce texts, following the guidelines shown in Table 5.
In all sessions, five experimental texts and five experiential texts were produced, which form the CAP. The order of elicitation of texts was the same in the three sessions.
The first session, in the informant's L1, included information about her linguistic background (following the format of the «European Language Portfolio»). The second session was developed in the informant's L2, and the third in English, with a part devoted to the linguistic ideologies of the informant. Each session also included a sequence of free conversation.

Lluís Payrató & Neus Nogué

The oral corpora in Catalan and research into casual Catalan:
the corpora of conversation and functional varieties of the CCCUB

Table 5. Session guidelines

			GENERAL DETAILS (LINGUISTIC BIOGRAPHY)		
1st day	**L1**	INTERVIEW	DIRECTED DISCOURSE	EXPERIMENTAL	Narrative	*Frog, where are you?*
					Instructive	Map of an imaginary village
					Argumentative	In favour: living in the country / city
					Descriptive	House from the SIMS computer game
					Expository	Immigration: the current situation
				EXPERIENTIAL	Argumentative	Recommend a film or book
					Instructive	From class to your home
					Descriptive	What your home is like
					Expository	What you do at the weekends
					Narrative	A dangerous or frightening situation
			CONVERSATION		
			EUROPEAN LANGUAGE PORTFOLIO	Language passport	
2n day	**L2**	INTERVIEW	DIRECTED DISCOURSE	EXPERIMENTAL	Narrative	*Frog, where are you?*
					Instructive	Map of an imaginary village
					Argumentative	In favour: living in the country / city
					Descriptive	House from the SIMS computer game
					Expository	Immigration: the current situation
				EXPERIENTIAL	Argumentative	Recommend a film or book
					Instructive	From class to your home
					Descriptive	What your home is like
					Expository	What you do at the weekends
					Narrative	A dangerous or frightening situation
			CONVERSATION		
3r day	**L3**	INTERVIEW	DIRECTED DISCOURSE	EXPERIMENTAL	Narrative	*Frog, where are you?*
					Instructive	Map of an imaginary village
					Argumentative	In favour: living in the country / city
					Descriptive	House from the SIMS computer game
					Expository	Immigration: the current situation
				EXPERIENTIAL	Argumentative	Recommend a film or book
					Instructive	From class to your home
					Descriptive	What your home is like
					Expository	What you do at the weekends
					Narrative	A dangerous or frightening situation
			CONVERSATION		
			Linguistic capital and ideologies		

Lluís Payrató & Neus Nogué
The oral corpora in Catalan and research into casual Catalan:
the corpora of conversation and functional varieties of the CCCUB

In the case of texts based on the experience of the speakers (*experiential texts*), the response was stimulated by a simple verbal request and no materials were used:

(a) Argumentative: recommending a film or book.
(b) Instructive: explaining how the informant goes from class to her home.
(c) Descriptive: describing the room and the apartment/house where the informant lives.
(d) Expository: the current situation of immigration in Europe.
(e) Narrative: narrating a dangerous or frightening situation.

In the case of eliciting experimental texts, responses were stimulated using visual materials (drawings and pictures), which guided the informants towards typical situations corresponding to:

(a) Narrative: a story about a frog that escapes and a boy who tries to find it.[7]
(b) Descriptive: the description of an apartment.
(c) Argumentative: preference for life in the country or life in the city.
(d) Expository: the current situation of immigration in Europe.
(e) Instructive: how to follow a route along the streets on a map of an imaginary village.

5.1.4 Formats and access. Examples

The CAP is conceived as a «monitor» corpus that can be expanded and which initially contains 360 core samples and 48 extra samples. In total the corpus contains approximately 18 hours of recordings (24 hours of interviews) and approximately 100,000 words in transcription. It is stored as a DVD (Payrató & Fitó 2008) and general tagging is not envisaged, being restricted to the adapted subcorpora.

The combination of text, sound and image means that the corpus contains text files (discourse transcription) with corresponding audiovisual recordings (image with sound). The following fragment is an example of the experimental narrative text (narration of a story prompted by images shown to the participant shortly before):

7. This well known story is based on the book by Mercer Meyer (1969): *Frog, where are you?*. New York: Dial Press.

Lluís Payrató & Neus Nogué
The oral corpora in Catalan and research into casual Catalan:
the corpora of conversation and functional varieties of the CCCUB

File C1TN21CS (fragment)

I21:   un nen que té una granota_
        en un potet de vidre_
        (..0,93) (INH) i::_
        m:_
        i bueno_
        i la deixen_
        (..0,59) que dormi\
        se suposa\
        que dormi\
        (..0,68) i que passi la nit allà amb ells\
        vale/
        i ell i el seu gos_
        perquè també té un gos_
        es posen a dormir\
        i al matí següent_
        quan es lleven_
        descobreixen_
        que la granota s'ha marxat_
        que no està dins el pot_
        i llavors comencen a buscar-la_
        comencen a crida:r-la_[8]

5.1.5 Corpus structure

The corpus structure is shown in the tables below, where samples are classified according to a combination of features relating to text type, genre, language (Catalan, Spanish, English), and elicitation procedures.

---

8. Literal translation: «A boy who has a frog in a small glass jar, and and well. and they let it sleep, it's supposed, sleep, and it spends the night there with them, OK?, and he and his dog, because he also has a dog, they go to sleep. And next morning, when they get up, they see that the frog isn't there, that it's not in the jar, and then they begin to look for it, they begin to call it.»

Lluís Payrató & Neus Nogué

The oral corpora in Catalan and research into casual Catalan:
the corpora of conversation and functional varieties of the CCCUB

Table 6. Corpus structure according to language and text type

CORPUS	TEXTUAL								
		Catalan	L1 or L2	experimental	Narrative	10	30 samples / informant	30 samples x 12 informants	= 360 samples
					Descriptive				
					Expository				
					Instructive				
					Argumentative				
				experiential	Narrative				
					Descriptive				
					Expository				
					Instructive				
					Argumentative				
		Spanish	L2 or L1	experimental	Narrative	10			
					Descriptive				
					Expository				
					Instructive				
					Argumentative				
				experiential	Narrative				
					Descriptive				
					Expository				
					Instructive				
					Argumentative				
		English	L3	experimental	Narrative	10			
					Descriptive				
					Expository				
					Instructive				
					Argumentative				
				experiential	Narrative				
					Descriptive				
					Expository				
					Instructive				
					Argumentative				

Lluís Payrató & Neus Nogué
The oral corpora in Catalan and research into casual Catalan:
the corpora of conversation and functional varieties of the CCCUB

Table 7. Corpus structure according to language and genre (conversation)

CORPUS	CONVERSATION	Catalan	L1	1	3 samples / informant	3 samples x 12 informants	= 36 samples
			L2				
		Spanish	L1	1			
			L2				
		English	L3	1			

Table 8. Corpus structure according to genre

CORPUS	TEXTUAL	30 +	34 samples x 12 informants	= 408 samples
	CONVERSATION	3 +		
	LANGUAGE IDEOLOGIES	1 sample / informant		
TOTAL		= 34 samples / informant		

5.1.6 Codification

The samples are coded according to four criteria: relation to the speaker, elicitation procedure, text type and informant number. The resulting code is shown as follows:

(1) Example of codification: C 1 T N 01 CC
(2) Relation to the speaker: 1 (first language) / 2 (second language) / 3 (third language)
(3) Elicitation strategy: T (experimental) / C (experiential) / E (interview)
(4) Type of text: N (narrative) / D (descriptive) / E (expository) / I (instructive) / A (argumentative) / C (conversation) / R (representations).
(5) Informant number: 01, 04, 05, 06 (L1 = Catalan) / 06, 12, 14, 15 (L2 = Spanish) / 21, 22, 24, 25 (L1 = Catalan/Spanish)
(6) Family language: CC (Catalan) / SS (Spanish) / CS (Catalan/Spanish) / EE (English) / DD (German).

Lluís Payrató & Neus Nogué
The oral corpora in Catalan and research into casual Catalan:
the corpora of conversation and functional varieties of the CCCUB

5.2. SAMPLES

In the following sections we provide a sample corresponding to each of the five main text types (all of them obtained as experiential texts regarding the elicitation strategy).

5.2.1 Narrative

C1CN04CC

1	I04:	vale\
2		una vegada vaig anar am una:_
3		discoteca_
4		(.. 0.56) i::_
5		(.. 0.45) feien una festa de l'espuma\
6		(.. 0.90) i::_
7		(.. 0.68) bueno_
8		(. 0.23) que--
9		que diverti:t_
10		l'espuma:_
11	ECC:	[mhm\]
12	I04:	[que guai_]
13		em vaig ficar al mig_
14		(.. 0.34) {(AC) és una discoteca que és com una piscina\
15		com una banyera molt gran\}
16		(.. 0.73) llavors va comença:r_
17		a pujar el nivell d'espuma_
18		fins que::_
19		(.. 0.31) i bueno:_
20		(. 0.16) fins aquí normal\
21		p(e)rò va arribar un moment que l'espuma em va cobrir\
22		(.. 0.64) i no podia respirar\
23		i no podia sortir\
24		i no sabia com fer-ho\
25		i x--
26		vaig estar a punt--
27		a punt d'ofegar-me allà\
28		(. 0.24) p(e)rò al final_
29		pues_
30		vaig poder sortir i:_
31		ja està\
32		(. 0.14) tot va quedar en això\
33		p(e)rò:_

Lluís Payrató & Neus Nogué

The oral corpora in Catalan and research into casual Catalan:
the corpora of conversation and functional varieties of the CCCUB

34		(.. 0.55) va haver un moment que_
35		vaig pensar_
36		o so- --
37		o:_
38		o aconsegueixo sortir ara_
39		o ja no surto d'aquí\
40	ECC:	(.. 0.40) mhm\
41	I04:	(.. 0.32) {(P) p(e)rò bueno\
42		{(??) (PP) vaig sortir\}}
43	ECC:	i va ser durant molta estona això o:_
44	I04:	home_
45		a mi se'm va fer molt llarg\
46		p(e)rò suposo que devia ser_
47		poc\
48		poc\
49		(.. 0.53) un minut com a molt\
50		perquè clar_
51		si t'ofegue:s_
52		(.. 0.35) vull di:r_
53		no pots aguantar molt estona\
54	ECC:	mhm\
55		(. 0.24) i:_
56		co:m_
57		com t'ho vas fer_
58	I04:	(.. 0.45) doncs em vaig puja:r_
59		(. 0.19) o sigui_
60		vaig_
61		enfilar-me a:l_
62		pòdium\
63		(. 0.21) hi havia un:_
64		pòdium_
65		que:_
66		envolta tota la:_
67		(.. 0.47) pista\
68		i jo estava al mig\
69		(.. 0.38) llavors clar_
70		tothom se'n va anar corrents cap a les bandes_
71		pero_
72		jo:_
73		pues_
74		em vaig queda:r_
75		(.. 0.62) am un pu:nt_
76		llavors vaig intentar pujar_

Lluís Payrató & Neus Nogué

The oral corpora in Catalan and research into casual Catalan:
the corpora of conversation and functional varieties of the CCCUB

77		vaig caure_
78		(.. 0.44) i llavors va ser_
79		va ser el moment de dir_
80		bueno\
81		o surto ara o no surto\
82		(.. 0.36) i vaig poder anar al final\
83		em van ajudar\
84		bueno\
85		quan van veure que estava allà em van:_
86		ajudar a pujar i ja està\
87	ECC:	(.. 0.48) mhm\
88	I04:	i vaig sortir\
89	ECC:	(. 0.16) molt bé\
90		o si[gui que--]
91	I04:	[i vaig] perdre una sabata\
92	ECC:	(.. 0.30) sí/
93		@
94	I04:	no\
95		p(e)rò després va aparèixer\
96		p(e)rò_
97		em pensava que me n'anava descalça\
98	ECC:	(.. 0.72) Déu n'hi do\
99		s'ha de passar malament_
100		no/
101		quan [estàs_]
102	I04:	[sí\]
103		sí\
104	ECC:	cobert_
105		fins aquí da:lt i:_
106		no pots sor[ti:r_
107		i tot_]
108	I04:	[sí sí\]
109		(. 0.19) clar_
110		és allò que:_
111		(.. 0.42) pensava que:_
112		[igual_]
113	ECC:	[i] tot per passar-t'ho bé\
114		no/
115	[	@@]
116	I04:	[sí sí\]

LLUÍS PAYRATÓ & NEUS NOGUÉ
The oral corpora in Catalan and research into casual Catalan:
the corpora of conversation and functional varieties of the CCCUB

5.2.2 Descriptive

C1CD22CS

1	I22:	e:l lloc/
2		{(??) vale\}
3	ECC:	sí\
4	I22:	(INH)
5		e::_
6		doncs bé\
7		(.. 0.80) m::_
8		(.. 0.32) le:s_
9		cases són:_
10		(.. 0.49) a ve(u)re\
11		hi ha_
12		hi ha de tots estils\
13		vale/
14		hi ha una mica d'antigues_
15		i així_
16		perquè:_
17		{(AC) la:_
18		zona no està malament\}
19		@@
20		[(INH)]
21	ECC:	[mhm\]
22	I22:	i::_
23		hi ha_
24		edificis_
25		no gaire alts\
26		tampoc\
27		de set o vuit pisos_
28		(INH)
29		i::_
30		le:s_
31		avingudes estan ben cuidades_
32		(INH)
33		[{(??) hi han moltes--}]
34	ECC:	[i el pis en concret_
35		una mica la_
36	I22:	[{(F) a:h\
37		el pis\}]
38	ECC:	[la distribu]ció_
39		{(??) sí\
40		co:m_}

Lluís Payrató & Neus Nogué

The oral corpora in Catalan and research into casual Catalan:
the corpora of conversation and functional varieties of the CCCUB

41	I22:	{(F) a:h\
42		[vale\]}
43	ECC:	[x]
44	I22:	(INH)
45		és gra:n_
46		@@
47		llavors té:_
48		(empassa saliva)
49		la cuina té un {(L2) office_}
50		és llarga\
51		(INH)
52		el menjador i e:l_
53		saló_
54		també són llargs\
55		hi ha un rebedor petit\
56	ECC:	mh[m\]
57	I22:	[lla]vors_
58		hi ha u:n_
59		{(AC) passadís_
60		que distribueix totes les habitacions_
61		que queden amb una banda\}
62		(INH)
63		i totes donen al carr:er\
64		llavors totes tenen:_
65		(INH)
66		tenen x una balconada_
67	ECC:	(. 0.20) [mhm\]
68	I22:	[que aga]fa tot el pis\
69		amb {(L2) ele\}
70		(INH)
71		i:_
72		llavors hi ha:_
73		dos ba- --
74		tres banys\
75		n'hi ha un de gran_
76		(INH)
77		i un de petitet al costa:t_
78		[(INH)]
79	ECC:	[mhm\]
80	I22:	i:_
81		n'hi ha un de:_
82		petit_
83		i:_

Lluís Payrató & Neus Nogué

The oral corpora in Catalan and research into casual Catalan:
the corpora of conversation and functional varieties of the CCCUB

84		i una habitacioneta_
85		que era pel servei_
86		que està al costat de la:_
87	ECC:	mhm\
88	I22:	de la cuina\
89		i que dóna a la galeria\
90		(.. 0.32) (INH)
91		i:_
92		teni:m_
93		(.. 0.52) (empassa saliva)
94		llavors tenim_
95		una {(P) habitació_
96		dos_
97		tres_}
98		quatre habitacions en total\
99	ECC:	(. 0.19) mhm\
100	I22:	(. 0.25) {(AC) i estan distribuïdes al llarg del passadís\
101		i al passadís hi ha:_}
102		(.. 0.85) hi ha uns arma:ris_
103		i_
104		{(P) de fet és això\
106		és com un rectangle\}
107	ECC:	mhm\

5.2.3 Expository

C1CE01CC

1	I01:	do:ncs:_
2.		(… 1.40) el divendres_
3		plego de treballar a les o:nze_
4		(.. 0.43) i m:_
5		(.. 0.58) me'n vaig a prendre algo\
6	ECC:	(.. 0.30) [{(P) mhm\}]
7	I01:	[però] també_
8		és que surto supercansada_
9		perquè m'he aixecat a les set {(@) del dematí_}
10		(.. 0.61) o sigui que:_
11		(.. 0.33) me'n vaig a dormi:r_
12		(.. 0.87) no sé_
13		cap a les du:gues o així\
14		(.. 0.74) {(AC) i el dissabte normalment em passo tot el dia al matí dormi:nt_}
15		(… 1.06) i_
16		per la tarda me'n vaig:--

Lluís Payrató & Neus Nogué

The oral corpora in Catalan and research into casual Catalan:
the corpora of conversation and functional varieties of the CCCUB

17		(. 0.17) cap allà a les ci:nc_
18		o (ai)xí_
19		vaig a fer un cafè_
20		(… 1.64) i_
21		si tinc que comprar alguna cosa_
22		o:_
23		qualsevol cosa_
24		(.. 0.54) i després a les vuit me'n vaig a treballar\
25		{(P) un atre cop\}
26		fins les onze\
27		@
28		(. 0.16) i el diumenge al dematí_
29		normalment_
30		em quedo mirant la tele pel dematí_
31		(.. 0.32) o em baixo al mercat de Sant Antoni a:_
32		(. 0.18) a buscar algo_
33		(.. 0.87) un llibre o_
34		alguna cosa\
35	ECC:	mhm\
36	IO1:	(.. 0.78) {(AC) i per la tarda me'n vaig a treballar {(@) una atra vegada\}}
37	ECC:	(.. 0.52) {(??) mira que bé\
38		[xx\]}
39	IO1:	{(@) [i] ja es[tà\]}
40	ECC:	[i] finalment_
41		quan arriba el diumenge_
42		[a les tardes_]
43	IO1:	[clar_
44		el diumenge a la nit_]
45		sí que em:_
46		normalment me'n vaig a sopar fora_
47		perquè és el únic dia_
48		així_
49		del cap de setmana_
50		(.. 0.48) que plego d'hora_
51		plego a les vuit i me'n vaig a sopar fora\
52	ECC:	(.. 0.30) mhm\
53	IO1:	(.. 0.47) {(P) i ja està\
54		cada setma[na és--]}
55	ECC:	[{(P) molt bé\}]
56		(. 0.26) cada cap de semana és ai[xí_igual_]
57	IO1:	[{(@) ts:\
58		sí:_}]
59	ECC:	sí_

Lluís Payrató & Neus Nogué

The oral corpora in Catalan and research into casual Catalan:
the corpora of conversation and functional varieties of the CCCUB

60	I01:	menos quan és algu:n_
61		així_
62		e--
63		(. 0.11) bueno\
64		algun dissabte_
65		pues me'n vaig de festa i tal_
66		{(P) però no:_}
67		(.. 0.81) normalment el cap de setmana és així\
68	ECC:	(. 0.28) si es treballa:_
69		[és el que hi ha_
70		no/]
71	I01:	[{(@) és xungo\
72		sí\}]
73	ECC:	és el que hi ha\

5.2.4 Instructive

C1CI01CC

1	ECC:	com s'hi va_
2		[què has de fer exactament_]
3	I01:	[doncs_
4		agafo:_]
5		la línia tres_
6		que la tinc just_
7		(. 0.20) just sortir de la universitat_
8		(. 0.11) creuar al mig de la Diagonal_
9		(.. 0.59) i agafo:_
10		la línia tres a Zona Universitària_
11		(.. 0.69) i baixo a plaça Espanya_
12		(.. 0.33) aleshores_
13		(.. 0.45) des de plaça Espanya vaig caminant a casa\
14		agafo el carrer Sepúlveda_
15		i tot recte\
16	ECC:	(.. 0.58) {(PP) molt bé\}
17		(.. 0.51) i molt fàcil_
18		[{(@) pel que veig\
19		no/}]
20	I01:	[{(@) sí\
21		(INH)}]
22		podria agafar la ver[mella_
23		{(??) que és una_

Lluís Payrató & Neus Nogué

The oral corpora in Catalan and research into casual Catalan:
the corpora of conversation and functional varieties of the CCCUB

24		parada_}]
25	ECC:	[(INH)
26		xxx el ca:- --
27		el carrer] Sepúlveda_
28		[i què_
29		i fins a on_
30		per (ai)xò_]
31	I01:	[@@
32		ah\
33		tot rec]te fins a::_
34		Entença\
35		{(P) o sigui_
36		pujant Entença\}
37	ECC:	mhm\
38	I01:	hi han cinc minutets\
39		{(??) p(e)rò_}
40		podria fer {(L2) transbordo_}
41		agafar la vermella_
42		que clar_
43		{(P) trigo mo:lt_
44		(.. 0.61) i em fa mandra\}
45	ECC:	(.. 0.46) {(??) et trobes} moltes vegades_ ((cop de porta))
46		camines més_
47		per sota [terra_
48		que] dic jo_
49	I01:	[mhm\]
50	ECC:	que no pas_
51		(. 0.12) per dalt no\
52	I01:	sí:\
53		(. 0.22) i:_
54		ho tinc molt {(@) a prop\}
55		(INH)
56	ECC:	molt bé\

5.2.5 Argumentative

C1CA04CC

1	I04:	do:ncs_
2		un llibre\
3	ECC:	(. 0.20) {(P) molt bé\}
4	I04:	(. 0.27) el llibre es diu:_
5		L'Alquimista_
6	ECC:	[{(P) mhm\}]

Lluís Payrató & Neus Nogué

The oral corpora in Catalan and research into casual Catalan:
the corpora of conversation and functional varieties of the CCCUB

7	Io4:	[i és de Paulo] Coelho\
8		(.. 0.99) i:_
9		és un llibre que::_
10		(. 0.18) sobre un noi_
12		que:_
13		fa un viatge_
14		que busca un tresor\
15		(.. 0.87) i::_
16		bueno\
17		la història_
18		vull di:r_
19		no:_
20		té:_
21		gra:n_
22		(.. 0.47) (fa espetegar la llengua)
23		(.. 0.61) no és que sigui una gran història_
24		allò:_
25	ECC:	mhm\
26	Io4:	(.. 0.44) p(e)rò:_
27		té u:n_
28		transfons_
29		vale/
30		molt_
31		(… 1.39) molt maco de::_
32		de una mena de filosofia_
33		o::_
34		(. 0.16) per dir-li d'alguna manera_
35		que:_
36		diu que::_
37		(.. 0.34) fa pensar que::_
38		(.. 0.91) que a la vida_
39		pues_
40		sempre pots aconseguir lo que tu vols_
41		i que::_
42		(.. 0.57) i que:_
43		tenir somnis_
44		doncs_
45		has de_
46		de lluitar pels teus somnis_
47		i::_
48		(.. 0.52) i que sempre_
49		doncs_
50		(.. 0.59) l'univers_

Lluís Payrató & Neus Nogué

The oral corpora in Catalan and research into casual Catalan:
the corpora of conversation and functional varieties of the CCCUB

51		t'ajudarà a conseguir lo que tu
		vulguis\
52	ECC:	(.. 0.41) mhm\
53	Io4:	i:_
54		{(P) jo què sé_}
55		t'acabes de llegir aquest llibre_
56		i:_
57		ets sents_
58		(.. 0.36) millor\
59		(… 0.99) {(P) és alegre i:_
60		(.. 0.46) mira\}
61	ECC:	(.. 0.64) molt bé\
62		(. 0.26) un llibre així_
63		optimi:sta[:_]
64	Io4:	[sí] [sí\
65		que:_]
66	ECC:	[reconfor]ta:nt_
67	Io4:	sí\
68	ECC:	(.. 0.64) {(P) molt bé\}
69		(.. 0.56) a l'època d'exàmens_
70		potser pot ajudar també_
71		llegir una cosa d'a[queste:s_
72		no/]
73	Io4:	[sí_
74		clar\]
75		també pot donar ànims\
76		p(e)rò bueno\
77		si tu_
78		en època d'exàmens_
79		comences a pensar que:_
80		tu lo que xx--
81		realment et vindria de gust_
82		és_
83		esta:r_
84		(.. 0.37) de viatge a::_
85	ECC:	[clar\]
86	Io4:	[la pla]tja:_
87		donc[s_]
88	ECC:	[x][xxxx] ((soroll de fons))
89	Io4:	{(somrient) [tampoc pot]ser no seria molt bo\}
90	ECC:	clar\
91		potser deixar-ho per després_
92		[no/]

Lluís Payrató & Neus Nogué
The oral corpora in Catalan and research into casual Catalan:
the corpora of conversation and functional varieties of the CCCUB

```
93   I04:      [sí\]
94   ECC:      (. 0.29) molt bé\
```

6. CLOSURE. A LOOK BACK AND FORTH AT THE SEARCH BASED ON ORAL CORPORA IN CATALAN

The Plurilingual Audiovisual Corpus (CAP) is the culmination of the project that began with the COC and COR corpora. It provides the first opportunity to analyze transcribed text and voice recordings together with the corresponding image, and therefore with several typical features of the multimodal communication that characterizes oral discourse. It also offers the first chance to study certain aspects of the Catalan spoken by the informants, together with phenomena related to their performancee of other languages (in particular, Spanish and English), taking into account the different types of text that the speakers usually construct: narrative, descriptive, expository, argumentative and instructive. In short, the CAP is best described as a multitextual, multimodal and multilingual corpus and we hope that these characteristics will foster progress in a field of research that has yet to be fully exploited.

In order to keep up with current debates on criteria for corpus construction (access, representativeness, marking, extension…), it is worth noting that the corpora presented here oriented towards qualitative research in pragmatics and discourse analysis (especially oral), without missing the possibility of quantifying data. The reason for such approach is that these are the work fields of the researchers who promoted the corpora, gathered under the heading GrEPAD (as stated in section 5).

Studies and discussions on the design of these corpora were born many years ago (see Payrató *et al.* 1996), with a title deliberately chosen to state the requirements and attributes of such collections and archives of texts, namely *Corpus, corpora*.[9] In turn, the areas of interest and research are also rooted in references from the last century, and they build on studies such as Payrató's (1988) seminal work on colloquial Catalan —among other's— which would lead to the COC, and on functional variation (Payrató 1998), the origin of the COR.

Later, analyses with a more markedly pragma-stylistic and multimodal orientation were also added to these trends (cf. Payrató 2003, 2018 and Payrató & Nogué 2013). Therefore, it is clear that the basic and general research carried out has always gone hand

---

9. This is the first volume —not by accident— of the so-called «Catalan Linguistic Collection» (see <http://www.publicacions.ub.edu/articulos.aspx?modo=c&fam=LINGÜÍSTICA CATALANA>).

Lluís Payrató & Neus Nogué
The oral corpora in Catalan and research into casual Catalan:
the corpora of conversation and functional varieties of the CCCUB

in hand with the elaboration of specific corpora. This explains the decision-making process carried out in their constitution, as well as their qualitative nature. Obviously, it should not be drawn from this that the corpora cannot be further improved and enriched, and that different decisions could not have been taken, nor that the research could not have been conducted in some way else.

From the current perspective it is worth pointing out a couple of remarks. Firstly, we know, at least from Romero-Trillo (2008), that pragmatics and corpus linguistics (cf. Habert *et al.* 1997; Biber *et al.* 1998, and many others) are «doomed» to understand in many aspects and levels. Secondly, that current general pragmatics must include, alongside *experimental pragmatics*, increasingly developed (see Noveck & Sperber 2004, inter alia), *corpus pragmatics*, as the journal *Corpus Pragmatics* has shown in a very clear way since 2017. Now all the old, aforementioned debates take on a new dimension, the relation between the concepts of theory, practice and application remaining the same.

The passing of the years also teaches us that corpora are useful when used. In fact, the three corpora described in this article have already been applied in several Ph. D. theses and other research projects. Texts from the COC were used in Castellà (2001 and 2004), Bladas (2000, 2003, 2006, 2009), Oller (2000), Payrató (1998), Payrató (2003, 2010), Nogué (2005, 2008a, 2008b, 2010), Matamala (2008), Cuenca & Torres (2008), Bladas & Nogué (2016), and De Cock & Nogué (2017). The COR was used in Nogué (2005, 2008a), and the CAP in Fitó (2009) and Lloberes & Payrató (2011). Additionally, all of these corpora have been used as data sources for multiple examples (see inter alia Payrató 2003, 2018).

Although Catalan still lacks an oral and written reference corpus providing comprehensive information about the language and allowing wide-ranging quantitative research, this study shows the value of the corpora presented here for conducting both quantitative and qualitative studies of many aspects of the structure and use of Catalan.

Lluís Payrató
Universitat de Barcelona
payrato@ub.edu
ORCID 0000-0001-8919-2886

Neus Nogué
Universitat de Barcelona
nnogue@ub.edu
ORCID 0000-0002-5671-3101

Lluís Payrató & Neus Nogué
The oral corpora in Catalan and research into casual Catalan:
the corpora of conversation and functional varieties of the CCCUB

BIBLIOGRAPHIC REFERENCES

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

Biber, Douglas, Susan Conrad & Randi Reppen (1998) *Corpus linguistics. Investigating language structure and use*, Cambridge, Cambridge University Press.

Bladas, Òscar (2000) «Les rutines de parla en el català col·loquial», in Vicent Salvador & Adolf Piquer (eds.): *El discurs prefabricat: estudis de fraseologia teòrica i aplicada*, Castelló de la Plana, Publicacions de la Universitat Jaume I, p. 325-331.

— (2003) *Les rutines de parla en el català col·loquial*, Ph. D. thesis, Barcelona, Universitat de Barcelona.

— (2006) «Com dir el que no sona. Les rutines de parla en discurs directe periodístic i col·loquial», in Vicent Salvador & Laia Climent (eds.), *El discurs prefabricat II: Fraseologia i comunicació social*, Castelló de la Plana, Universitat Jaume I, p. 267-301.

— (2009) *Manual de transcripció del discurs oral. Materials de treball*. Barcelona: Universitat de Barcelona / PPU. [On line: <http://diposit.ub.edu/ dspace/ bitstream/2445/106301/1/Manual de transcripció.pdf>.]

Bladas, Òscar & Neus Nogué (2016) «"Que bé, *tu*!" ("That's great, *you*"): an emerging emphatic use of the second person singular pronoun *tu* (*you*) in spoken Catalan», *Pragmatics*, 26, p. 473-500.

Boix-Fuster, Emili, Marina Àlamo, Mireia Galindo & F. Xavier Vila (2006) *Corpus de varietats socials. Materials de treball*, Barcelona, Publicacions de la Universitat de Barcelona.

Castellà, Josep M. (2001) *La complexitat lingüística en el discurs oral i escrit: densitat lèxica, composició oracional i connexió textual*, Ph. D. thesis, Barcelona, Universitat de Barcelona.

— (2004) *Oralitat i escriptura. Dues cares de la complexitat del llenguatge*, Barcelona, Curial Edicions Catalanes / Publicacions de l'Abadia de Montserrat.

Cuenca, M. Josep (2011) «Catalan interjections», in Lluís Payrató & Josep M. Cots (eds.), *The Pragmatics of Catalan*, Berlin/Boston, De Gruyter Mouton, p. 173-211.

Cuenca, M. Josep & Marta Torres (2008) «Usos de *hombre/home* y *mujer/dona* como marcadores del discurso en la conversación coloquial», *Verba*, 35, p. 235-256.

De Cock, Barbara & Neus Nogué (2017) «The pragmatics of person reference: A comparative study of Catalan and Spanish parliamentary discourse», *Languages in Contrast*, 17, p. 96-127.

Lluís Payrató & Neus Nogué
The oral corpora in Catalan and research into casual Catalan:
the corpora of conversation and functional varieties of the CCCUB

Du Bois, John W., Stephan Schuetze-Coburn, Susanna Cumming & Danae Paolino (1993) «Outline of discourse transcription», in Jane A. Edwards & Martin D. Lampert (eds.), *Talking data: Transcription and coding in discourse research*, Hillsdale, Lawrence Erlbaum, p. 45-89.

Habert, Benoît, Adeline Nazarenko & André Salem (1997) *Les linguistiques de corpus*, Paris, Colin - Masson.

Lloberes, Marina & Lluís Payrató (2011) «Pragmatic coherence as a multimodal feature: Illustrative cospeech gestures, events, and states», in Lluís Payrató & Josep M. Cots (eds.), *The Pragmatics of Catalan*, Berlin/Boston, De Gruyter Mouton, p. 215-246.

Matamala, Anna (2008) «La oralidad en la ficción televisiva: análisis de las interjecciones de un corpus de comedias de situación originales y dobladas», in Jenny Brumme (ed.), *La oralidad fingida: descripción y traducción*, Madrid, Iberoamericana - Vervuert.

Nogué, Neus (2005) *Dixi de persona i marcs participatius en català*, Ph. D. Thesis, Barcelona, Universitat de Barcelona / Departament de Filologia Catalana.

— (2008a) *La dixi de persona en català*, Barcelona, Publicacions de l'Abadia de Montserrat.

— (2008b) «La dixi de persona en el discurs acadèmic oral en català», *Caplletra. Revista Internacional de Filologia*, 44, p. 195-218.

— (2010) «La primera persona del plural en català», *Llengua & Literatura*, 21, p. 155-198.

Noveck, Ira A. & Dan Sperber, eds. (2004) *Experimental pragmatics*, New York, Palgrave Macmillan.

Oller, Anna (2000) «Connectors i fraseologia: el *llavors* col·loquial», in Vicent Salvador & Adolf Piquer (eds.), *El discurs prefabricat: estudis de fraseologia teòrica i aplicada*, Castelló de la Plana, Publicacions de la Universitat Jaume I, p. 333-342.

Payrató, Lluís (1995) «Transcripción del discurso coloquial», in Luis Cortés (ed.), *El español coloquial*, Almería, Universidad de Almería, p. 45-70.

— (2003) *Pragmàtica, discurs i llengua oral*, Barcelona, UOC (2nd ed., 2010).

— (2018) *Introducción a la pragmática. Una perspectiva sobre el lenguaje en acción*, Madrid, Síntesis.

—, ed. (1998) *Oralment. Estudis de variació funcional*, Barcelona, Universitat de Barcelona / Publicacions de l'Abadia de Montserrat.

Payrató, Lluís & Núria Alturo, eds. (2002) *Corpus oral de conversa col·loquial. Materials de treball*, Barcelona, Publicacions de la Universitat de Barcelona.

Payrató, Lluís, Emili Boix, M. Rosa Lloret & Mercè Lorente, eds. (1996) *Corpus, corpora*, Barcelona, PPU / Universitat de Barcelona.

Lluís Payrató & Neus Nogué
*The oral corpora in Catalan and research into casual Catalan:
the corpora of conversation and functional varieties of the CCCUB*

Payrató, Lluís & Jaume Fitó, eds. (2008) *Corpus audiovisual plurilingüe*, Barcelona, Universitat de Barcelona.

Romero-Trillo, Jesús, ed. (2008) *Pragmatics and corpus linguistics. A mutualistic entente*, Berlin, Mouton de Gruyter.

Subirats, Marina, Marta Massats & Cristina Sánchez (1992) «11. Els hàbits lingüístics», in Marina Subirats (dir.), *Enquesta de la Regió Metropolitana de Barcelona 1990*, Barcelona, Institut d'Estudis Metropolitans de Barcelona.