

A pesar de estas y otras objeciones, el volumen representa un esfuerzo importante de estos investigadores, que, quizás, pueda mejorarse en estos y otros aspectos para la edición definitiva de los materiales. Se trata de un libro muy atractivo, con excelente impresión pero con algunos detalles pequeños de cuidado de la edición que también podrían ser mejorados. Por ejemplo, la numeración no continua de los mapas del volumen también pueden dar lugar a confusiones.

*Adolfo Elizaincín*

*Universidad de la República, Montevideo*

**Lluís Payrató/Núria Altura (eds.): *Corpus oral de conversa col·loquial. Materials de treball*. Barcelona: Publicacions de la Universitat de Barcelona (Universitat; 11) 2002. 98 pàgines + 1 CD-ROM. ISBN 84-475-2597-X.**

The present booklet and the CD-ROM that comes with it constitute the first publicly available output of a large corpus-linguistic project started by the Department of Catalan Philology of Barcelona University in the early 1990's, the "Corpus de Català Contemporani de la Universitat de Barcelona" (CUB). The CUB project includes a total of 7 subcorpora, all except one of them focussing on spoken language and ranging from a more traditional corpus of dialect samples in different Catalan varieties ("Corpus Oral Dialectal", COD) to rather specific corpora documenting, e.g., news broadcasting speech ("Corpus d'Informatius Orals", CIO) or radio advertisements ("Corpus Oral de Publicitat", COP) (for a general presentation of the CUB project cf. Alturo /Boix/Perea 2002). Among the seven components of CUB, a subgroup of three corpora – among them the corpus

under review – may be distinguished on the basis of certain common features, i.e. the fact that they document primarily the Central Catalan variety spoken in the Barcelona metropolitan area, and their focus on spontaneous or near-spontaneous speech. Apart from the "Corpus Oral de Conversa Col·loquial" (COC), this subgroup is made up of the "Corpus Oral de Varietats Socioculturals" (COS), an interview corpus based on a balanced sample of 78 informants selected according to basic socio-linguistic variables; and the "Corpus Oral de Registres" (COR), a corpus of spoken interaction in a large variety of situational contexts and social domains. Although no definite figures are given, each of these three spontaneous spoken language corpora seems to contain about 350.000 words, thus yielding a mid-size corpus of slightly more than a million tokens. As for the COC sub-corpus, its total extension is calculated at 357.500 words; the sample published in the book and on the CD-ROM under review contains some 70.500 words. This seems to be few data, but constitutes nonetheless the quantitatively most important corpus of spontaneous spoken Catalan published hitherto (for some more publicly available corpora of spoken Catalan cf. Pusch 2002: 258).

The slim booklet consists of an introductory chapter and some 70 pages of transcriptions. The introduction is a modified version of a paper originally published as Oller et al. (2000) and gives a general overview of the COC and more detailed information on the 10 selected texts (out of 50 which constitute the entire corpus) transcribed and recorded in the book and on its companion CD-ROM. Apart from technical descriptions, this introduction contains a list with basic biographical data on the 58 informants documented in the sample, data which is complemented by some socio-linguistic vari-

ables such as level of education, socio-economic status, and explicit knowledge of Standard Catalan. Furthermore, the transcription conventions are explained and an exhaustive list of the collaborators of the CUB project is given. In the second part of the booklet, the first 250 to 350 turns or intonation units of each of the 10 texts are given in both orthographic ("discursive") and phonetic transcription. It therefore becomes obvious that the real work-tool of this publication is the CD-ROM, to which the booklet constitutes a mere annex.

The CD-ROM contains the full orthographic transcriptions of the 10 texts (but still the partial phonetic transcriptions as they appear in the booklet only) and the corresponding recordings. The whole architecture of the CD-ROM is HTML-based and therefore requires no previous installation (and subsequent uninstall procedure) to use the transcriptions. Once introduced into the CD-ROM device, the user only has to open the COC.html file in a current browser to visualize the corpus' introductory page, from which one is conveniently guided (albeit in Catalan only; just a short presentational text is available in Spanish and English) through the different parts of the program. The decision of the authors to conceive their product in 'orthodox' HTML certainly limits the possibility of integrating interactive features but has the huge advantage of making sure that the CD-ROM will run perfectly on many machines and still will do so in a couple of years, when many of today's more sophisticated multimedia products will get virtually inaccessible due to emerging incompatibilities of the software, or software versions, they were created with. The audio files, however, are stored on the CD-ROM as Shockwave Flash objects, so that the user of the CD-ROM has to have the Macromedia Flash

player installed and the corresponding plug-in implemented with his browser to correctly access the recordings from the transcription screens. It might have been a good idea to include the audio files in an additional folder in the more neutral MPEG-3 format, but users with some experience in computer-based audio handling will easily create these MP3 files themselves.

The 10 texts contained in this corpus sample are accessed from an index page that gives access to the printable versions of the transcriptions (which come in Microsoft WORD format) and to the meta-informations (transcription conventions etc.) already described in the book. The transcription pages contain the same links and additional linked documents, such as contextual information on the situation documented in the transcription, and data on the informants who participate in the current conversation. As a matter of fact (and in accordance with the title of the corpus), all of the 10 texts of this sample are conversational, with up to 11 speakers involved in the interaction. The recordings have generally been done in surroundings very familiar to all the informants and, in many cases, in family homes with family members and intimates discussing in a casual and informal way, thus making sure that a high degree of spontaneity and authenticity of the documented conversations is attained. Spontaneity and informality is enhanced by the fact that certain informants had initially been recorded secretly, with authorization to use the recordings being obtained later. As far as the author of these lines has been able to evaluate these parameters, the 10 sample texts of the COC are indeed highly authentic and representative of relaxed spoken discourse, the conversations evolving around typical subjects of daily life in mostly private contexts.

The transcriptions appear on the screen in portions of approximately 40 lines per page/screen. The authors have chosen a linear notation, with each turn or – in the case of longer turns by the same speaker – each intonation unit, as well as exceptionally long breaks, being attributed a new transcription line. They have followed the discourse transcription system proposed by J. W. Dubois et al. (1991), which apart from breaks and some extra-linguistic features such as laughing allows for the notation of various prosodic features, such as rising or falling tones, voice intensity, or allegro and lento speech. All these additional features are noted within the transcribed speech and separated from it by rounded or curved brackets. Obviously, such a linear notation system does not allow for iconic representation of overlapping or simultaneous speech events, which are marked by square brackets instead. This is not so much of a disadvantage, as the present corpus allows to listen to the recordings while reading the transcriptions, and to easily locate and reconstruct such cases of simultaneous speech, if this is the guiding interest of the corpus user. The recordings are cut into pieces according to the transcription screens (with approximately 40 transcription lines each, as explained before), and by clicking from screen to screen, the user may follow the entire conversations, which are of about 30 minutes each. However, text and sound are not fully aligned, and there is no visual element to indicate the progress of the recording within the text (contrary to the recently published Portuguese corpus edited by Bacelar do Nascimento 2001). The sound quality of these audio recordings is highly variable and, in some cases, rather poor, so that for a non-native speaker of Catalan, some recordings remain precarious if not unin-

telligible if not supplemented by the transcription. This is obviously the price to be paid for 'authentic' spoken-language data probably recorded with unobtrusive recording devices and hidden microphones, a fact that will certainly preclude the use of the recordings for any in-depth phonetic analysis, but one has to bear in mind that this is not the primary goal of the corpus, which, according to the explanations given by its authors in the booklet (p. 8), is intended primarily as a tool and data base for pragmatically oriented research ("estudis de descripció i anàlisi del català col·loquial des d'una perspectiva bàsicament pragmàtica (anàlisi conversacional, fenòmens textuais i gramaticals, etc.).").

As mentioned above, the CD-ROM offers at various steps a download option for both the orthographic and the phonetic transcriptions. As these download files are in MS WORD format, the user will have to convert them to ASCII format in order to analyze them with the current corpus-linguistic tools such as concordancers or frequency calculators. Unfortunately, the transcription system applied by the authors, which makes use of various types of brackets but does not include the square brackets used, for instance, for HTML tags, does not allow to easily strip off the paralinguistic parts of the transcription and to fade out everything except for the linguistic production of the informants, a feature that might be welcome for studying, e.g., collocations on the basis of concordances. Again, this is not too much of a disadvantage, as it remains questionable if a small corpus such as this COC sample presents itself for advanced quantitative analysis, for which these corpus-linguistic tools are designed and normally used.

All in all, the present publication makes available to scholars of present-day

Catalan highly useful materials presented in a simple and efficient way and allowing for convenient application. Obviously, quantitatively speaking, contemporary oral Catalan is far behind Spanish or Portuguese, as far as available corpora are concerned, but with the present booklet and CD-ROM, a first step is made and one would wish to see in not too far a future the remainder of the COC data and other components of the CUB projects published in a similar manner.

#### References:

- Alturo, Núria/Boix, Emili/Perea, M. Pilar (2002): "Corpus del Català Contemporani de la Universitat de Barcelona (CUB): a general presentation", in: Pusch, Claus/Raible, Wolfgang (Ed.): *Romanistische Korpuslinguistik. Korpora und gesprochene Sprache/Romance corpus linguistics. Corpora and spoken language*. Tübingen: Narr (ScriptOralia; 126), 155-170.
- Bachelor do Nascimento, Maria Fernanda (ed.) (2001): *Português falado. Documentos autênticos. Gravações audio com transcrição alinhada*. Lisboa: Centro de Linguística da Universidade de Lisboa/ Instituto Camões (on CD-ROM).
- Dubois, John W. et al. (1991): *Discourse transcription*. London: Longman.
- Oller, Anna et al. (2000): "El COC del CUB: un corpus per a l'estudi de la conversa col·loquial", in: *Zeitschrift für Katalanistik* 13, 58-91.
- Pusch, Claus D. (2002): "A survey of spoken language corpora in Romance", in: Pusch, Claus/Raible, Wolfgang (Ed.): *Romanistische Korpuslinguistik. Korpora und gesprochene Sprache/Romance corpus linguistics. Corpora and spoken language*. Tübingen: Narr (ScriptOralia; 126), 245-264.

Claus D. Pusch  
Albert-Ludwigs-Universität Freiburg

**Graciela Reyes: *Metapragmática: Lengua sobre lenguaje, ficciones, figuras*. Valladolid: Universidad de Valladolid (Linguística y filología, 50) 2002. 116 páginas. ISBN 84-8448-184-0.**

#### **Metapragmática: ¿una nueva perspectiva dentro del estudio de la pragmática y la lengua?**

La metapragmática estudia el compendio de conocimiento que hace posible la competencia de los hablantes. Incluye el conocimiento del lenguaje, del mundo social, de los parámetros de uso del lenguaje, de las normas de comportamiento, de las normas lingüísticas, y de lo relacionado con los textos y su papel en nuestra vida. El principal interés de un análisis desde el punto de vista metapragmático, como es el llevado a cabo en *Metapragmática: Lengua sobre lenguaje, ficciones, figuras*, es proporcionar una nueva perspectiva dentro del estudio de la pragmática y la lengua. En esta obra, Reyes proporciona una introducción a los conceptos básicos de la pragmática e intenta introducir al lector tanto en conceptos teóricos fundamentales como en la bibliografía básica, sin perder de vista una perspectiva lingüística más profunda y sin rehuir tampoco la profundización en los temas más actuales y problemáticos, puesto que ofrece una revisión de las tendencias teóricas y de las investigaciones más recientes en los ámbitos de la relevancia, la voz y la ironía.

Mediante la nueva perspectiva de análisis denominada metapragmática se busca averiguar cómo la producción de la lengua está íntimamente ligada al conocimiento del individuo. La metapragmática se puede definir como "la competencia de los hablantes sobre el uso de la lengua" (p. 23) y también como "una normativa, a veces implícita, que refleja las ideas que