Review

Parodi, Giovanni (ed.) (2007): *Working with Spanish Corpora*.
London/New York: continuum.

Sönke Matthiessen (Frankfurt/Oder)

In today's linguistic research, there are still few publications in the growing field of corpus linguistics. Due to the comparatively small number of studies conducted in this discipline, there is only a handful of works that concentrate on Spanish. This is quite astonishing, given that Spanish is one of the most-spoken languages. Therefore, *Giovanni Parodi* and his fellow authors make the effort to at least "partially fill this gap by organizing a collection of investigations following the principles of corpus linguistics" (Parodi 2007: 1), that investigate contemporary Spanish in Spain and the Americas; namely Argentina, Chile, the U.S. and Venezuela. While all contributors to this volume use different approaches and methodology, they all do share "the objective of describing linguistic structures of Spanish from a corpus linguistics perspective" (ibid: 2).

The book is divided into ten chapters by just as many contributors. Parodi himself and three of the other authors are involved with the influential *Pontificia Universidad Católica de Valparaíso* (Chile). The others come from several distinguished institutions such as the *Universidad de Buenos Aires* (the largest university in Argentina), the *Iowa State University* (member of the prominent AAU) or the still very young but already quite prestigious *Universitat Pompeu Fabra* of Barcelona.

With regards to contents, the common ground for this compilation of essays lies in their use of spoken and written corpora. More precisely it can be stated, that the emphasis of all articles is put on the exploration of linguistic variation in Spanish while different registers (spoken and written, specialised and non-specialised) are examined. This makes Parodi's book intriguing for both English-speaking linguists interested in Spanish and researchers of contrastive linguistics.

Due to the limited space, the articles of this ground-breaking work cannot be discussed in depth; therefore their presentation will have to be limited to just a few short summaries of the most intriguing ones.

Following a brief but concise introduction (Chapter 1), Parodi provides a selection of resources for the corpus linguistic analysis of contemporary Spanish as well as some other languages. The list offers a wide range of corpora, websites on research programs and links to computational resources, all of which are supplied with a brief description and/or comment, that may serve as a starting point or at least a small impulse for future scientific projects (although not necessarily indicated as such by Parodi).

In Chapter 2, as well written by the editor himself, Parodi tries to scrutinize linguistic variation across registers in Spanish with a corpus linguistics point of departure. In order to do so, the author analyzes the *El Grial PUCV-2003 Corpus*, which consists of 90 texts (a total of more than 1.4 million words) trying to statistically determine "the salient linguistic and co-occurring patterns" (ibid: 15).

Chapter 3 by *Douglas Biber* and *Nicole Tracy-Ventura*, both from *Arizona Northern University*, is dedicated to the dimensions of register variation in Spanish. In the following chapter (4), *Guiomar E. Ciapuscio* presents us with a pilot study for a new corpus based on

texts "of scientific Spanish used in Argentina, both in oral and written varieties" (ibid: 91) which is to be called COTECA (*Corpus Textual del Español Científíco de la Argentina*). The design of the corpus provides a promising and refreshingly new approach to scientific discourse, given that "the traditional attitude in linguistics [...] considers the written and monolingual variety to be the exemplary model" (ibid: 93), neglecting the study of oral academic discourse, which is exactly what the COTECA-project is taking into account. The aims are to broaden the "knowledge of lexical, grammatical and textual topics" as well as to gather "descriptive knowledge that will contribute to the education and training of young scientists and specialists" (ibid: 93f.).

Omar Sabaj provides us with a multi-register analysis of prepositional schemes in communication verbs in Spanish (Chapter 5), while *Mercedes Sedano* of the *Universidad Central de Venezuela* presents a research on the spoken and written varieties of future tense expressions in several Spanish corpora in Chapter 6. Chapter 7, again by the very diligent *Giovanni Parodi* and *Aída Gramajo*, deals with technical-professional discourses.

In the subsequent chapter (8), entitled "Academic writing: Exploring Corpus 92", *Carmen López Ferrero* addresses the problem of the "shortage of studies on learner corpora [...] to explain the level of communicative competence [...] [and] suggest the appropriate teaching method" (ibid: 173). The design of Corpus 92, which has been included in the CREA of the RAE, gives access to reliable information "on the degree of written competence possessed by students upon entering university" (ibid.). The objective of the analysis of such a corpus is self-explanatory; it serves as an implement in designing a valid basic concept which provides efficient teaching materials that will meet the needs of the students.

Chapter 9 ("Using Latent Semantic Analysis in a Spanish research article corpus") by *René Venegas* and Chapter 10, dedicated to lexical bundles in speech and writing, by *Nicole Tracy-Ventura*, *Viviana Cortes* and *Douglas Biber* complete this excellent opus, that although "aimed primarily at English-speaking linguists" (ibid: 2) may also be very valuable to undergraduate and postgraduate students interested in the field of corpus linguistics.