
**THOMAS SCHMIDT AND KAI WÖRNER (EDS.) 2012.
MULTILINGUAL CORPORA AND MULTILINGUAL CORPUS
ANALYSIS. ÁMSTERDAM: JOHN BENJAMINS
PUBLISHING, 407 PAGES. ISBN: 978-90-272-1934 3.**

Reviewed by Cristina Toledo

University of Alcalá, Spain
cristina.toledo@uah.es

Multilingual Corpora and Multilingual Corpus Analysis focuses on linguistic aspects of multilingualism, specifically on the design of corpora in studies on multilingualism as well as presentations of linguistic analyses conducted using multilingual corpora. It is worth mentioning that the term “multilingual corpus” is broader than the commonly accepted definition (parallel corpora with translations of the same texts into different languages). By multilingual corpus the editors refer to “any systematic collection of empirical language data which enables linguists to carry out analyses about multilingual individuals, multilingual societies or multilingual communication” (p. xi).

The volume is a collection of 22 contributions, 13 presenting studies conducted at the Research Center on Multilingualism at the University of Hamburg (Germany) and nine describing external projects. Starting with a brief editors’ introduction, the book is organized in five sections: 1) learner and attrition corpora; 2) language contact corpora; 3) interpreting corpora; 4) comparable and parallel corpora; and 5) corpus tools. However, the distribution of the papers is not exactly balanced: some sections have more than five contributions whereas others have only two. The volume concludes with a general index, a corpora index and a language index.

Section 1. Learner and Attrition Corpora

Section 1 is the largest, with nine studies, and it explores the creation and analysis of various multilingual learner corpora of different sizes.

The first four studies in this section focus on learner corpora containing data from adult learners. Gut’s opening chapter presents the Learning Prosody in a Foreign Language corpus of spoken learner German and learner English, which includes spoken samples produced by both native and non-native speakers of two languages in four different speaking styles. Focusing on the reusability of a corpus, in the second chapter, Hedeland and Schmidt report on methodological problems connected with the creation and annotation of a spoken German corpus, arguing that reliable annotations are of paramount importance for corpus reusability. Ott, Ziai and Meurers, in the following chapter, also focus on the annotation procedure and inter-annotator agreement to a written Corpus of Reading Comprehension Exercises in German. They explore the appropriateness level of answers of adult learners of German to reading questions and the results show most answers are appropriate when taking the meaning assessment. In the next paper, Zinsmeister and Breckle present a text-based corpus of two sub-corpora of argumentative essays written by advanced learners of English and the authors offer suggestions for the use of this specific corpus when teaching German.

The next three chapters describe corpora containing data produced by young learners. In the fifth contribution by Saceda Ulloa, Lleó and García Sánchez, recordings of bilingual

speakers of Spanish and German are compared with those of monolingual German children in terms of the characteristics of their spoken language. In the sixth chapter, Lleó describes two corpora —one from German and Spanish monolingual children and another produced by German-Spanish bilingual children— created over 25 years with the purpose of investigating phonological first language acquisition of German-Spanish bilingual children. In chapter seven, Herkenrath and Rehbein present a bilingual Turkish-German and a monolingual Turkish corpus of spoken child language and apply a methodology called Pragmatic Corpus Analysis.

The final two chapters in Section 1 focus on attrition corpora. Czachór reports on a Polish-German bilingual written and spoken corpus with the purpose of exposing change on morphosyntactic features. The final paper of this section, by Kupisch, Barton, Bianchi and Stangen, deals with a corpus of German-French and German-Italian bilingual adults in order to show acquisition deficits at linguistic domains.

Section 2. Language Contact Corpora

In Section 2, the contributions provide an overview of language contact corpora. These include two main types: corpora exploring varieties of language that have developed, and continue to do so, in a language contact context —four first contributions—; and corpora examining the establishment of a language which developed through a history of language contact —fifth contribution—. The first contribution, by Gabriel, addresses the impact of migration contact with Italian and its dialects on two varieties of the Argentinian-Spanish prosodic system. In the second chapter, Köhl presents a corpus containing spoken and written data in Faroese, Danish and Faroese Danish and reports that language contact effects depend on two factors: the medium (spoken or written) and the register (formal or informal).

In the following paper, Benet, Cortés and Lleó use a spoken corpus of Catalan to investigate particular phonological aspects of Catalan found in bilingual Catalan-Spanish speakers from three different age groups and from two different areas of Barcelona. Putz's paper is based on a corpus of medical dialogues among L1 Italian doctors and L1 German patients in Tyrol in order to detect the elements causing communication obstacles when communicating in German. In the final chapter of the section, Höder examines the development of syntactic change in Old Swedish, a language highly influenced by language contact.

Section 3. Interpreting Corpora

Section 3 is the most relevant and interesting for public service interpreting because it provides three contributions from Interpreting Studies, revealing some interesting insights into the use of corpora, particularly in public service interpreting settings, and on the need for more interpreting corpora and corpus-based approaches in interpreting research.

Angermeyer, Meyer and Schmidt highlight the need to create and share data from community interpreting corpora to explore linguistic aspects of interpreting and also how interpreters deal with complicated communicative situations. In the subsequent paper, Bührig, Kliche, Meyer and Pawlack present how to communicate messages in a medical environment and address how data from the analysis of the corpora containing these messages can have a training purpose. Finally, Bührig, Kliche, Meyer and Pawlack present a brief description of a corpus of ad-hoc-interpreting in German hospitals intended to be used for training in bilingual workplace communication.

Section 4. Comparable and Parallel Corpora

In this section, two types of corpora, used primarily for Translation Studies, are included: corpora containing texts produced in similar settings but with different languages or

language varieties and parallel corpora in which original texts are aligned with their translations into other languages.

The first paper by Fandrych, Meißner and Slavcheva describes a parallel spoken academic corpus from German, English and Polish concentrating on two academic genres (presentations and academic papers). As the authors point out, this is the first corpus of spoken academic German that can be used for contrastive investigations. Dittman, Āurćo, Geyken, Roth and Zimmer, in the second paper, present a written corpus of German varieties, entitled Korpus C4, with the purpose of tracing the use of the German language throughout the 20th and 21st century in Germany, Switzerland, Austria and Tyrol, Italy. The potential uses of this corpus include research into differences between the varieties, issues related to language contact, and pedagogical applications as well as uses in language teaching and learning.

In the final chapter of this section, Āulo and Hansen-Schirra present a parallel corpus of German and English source texts and their translations, and argue for the use of dependency treebanks in Translation Studies research.

Section 5. Corpus tools

The last section in the volume focuses on some practical tools that corpus linguists might use when creating and analyzing multilingual corpora. The section has only two chapters. In the first chapter, Rose describes the PhonBank tool, which has been developed within the framework of the CHILDES project. This free and open-source tool brings new functions to the corpus building and analysis. In the final contribution, Wćrner explains how metadata are handled by the EXMARaLDA (Extensible Markup Language for Discourse Annotation) system and concludes with comments about standardizing metadata.

To summarize, *Multilingual Corpora and Multilingual Corpus Analysis* reflects the diversity in multilingual corpus studies. Consequently, albeit mainly targeted at corpus linguists and corpus researchers, the volume might also be of interest to language scholars as it provides empirical data and tools for their research. As far as public service interpreting and translation are concerned, Section 3 is the most relevant, even though other sections also offer useful insights into pedagogical implications of multilingual corpora that can be applied to translation and interpreting training.