

**Margaret Deuchar, Peredur Webb-Davies, Kevin Donnelly:  
Building and Using the Siarad Corpus. Bilingual conversations in  
Welsh and English**

(John Benjamins Publishing Company, Amsterdam/Philadelphia. 2018. 199 p.)

"Building and Using the Siarad Corpus. Bilingual conversations in Welsh and English" presents a detailed analysis of code-switching patterns in 40 hours of spontaneous conversational data from 151 bilingual Welsh-English participants. The main aim of the book is to show how such large data sets can inform us about code-switching patterns within and between bilingual speakers.

The book is organized in two parts: "*Building the corpus*" and "*Using the corpus*". The first part consists of three chapters on the process of setting up the corpus. The second part consists of five chapters, and the main focus of this section is the actual hands-on work with the Siarad corpus. Appendixes contain documentation for the Siarad corpus and questionnaires for participants, a list of references, and an index. Besides, each chapter ends with a short paragraph, summarizing the most important points of that chapter.

The first chapter is the introduction, where authors describe the main ideas of their work and raise 4 important theoretical questions which are essential for reading the book:

*What is a bilingual speech corpus and how it is useful?*

*What is code-switching?*

*What is the historical context of the Siarad corpus?*

*What is the Siarad corpus?*

This part gives the reader a proper insight into the theoretical base of the research, which makes the book accessible even for students and unexperienced researchers in this field of linguistics.

The second chapter opens the section titled "*Building the corpus*". The chapter contains the main information about the quality and type of the collected data, methods of data collection, and main characteristics of speaker profiles. The Siarad corpus presents data in the form of transcripts of conversations between two participants rather than a group of speakers.

The participants were English and Welsh bilinguals, who were recruited by letters in Welsh and English sent by the researchers in the local area (Bangor University, Wales). Of the recruitment methods used by the researchers for the target audience of Bangor University students and people outside the university, the social network and "friend of a friend" approach turned out to be more successful than posters and advertisements. The conversations between the participants were recorded on audio, without any presence or interference of the researchers to achieve a high degree of

naturalness and adequacy. After the recording the participants completed background questionnaires of 20 questions about their age, gender, occupation, geographical area, educational background, age of language acquisition, self-reported level of proficiency in both languages, language input at home, the language of education, languages of social networking, attitudes to language, self-reported identity, self-reported use of code-switching and the attitude to code-switching. The description of each part of the questionnaire contains visual materials in the form of graphs and tables, which show a wide range of subjects according to their linguistic background.

Chapter 3 covers various aspects of the data transcription: the choice of the transcription system (CHAT), the process of linking transcriptions to the sound, the use of an autoglossing system (CLAN), and also reliability and availability issues of the whole corpus. In this chapter, the reader can see how transcribed information is represented in the Siarad file.

Chapter 4 is concerned with the differences between code-switches and borrowings. In the chapter the authors present two opposing points of view on this issue (Myers-Scotton vs Poplack & Meechan) and describe the main differences between them. According to the authors, code-switches have a lower level of frequency and integration than borrowings.

The part on "*Using the Corpus*" opens with chapter 5, dedicated to the morphosyntactic patterns Welsh-English bilinguals use when they code-switch. The chapter is focused on the matrix language concept. Here the authors mention "the switching rule" introduced by Joshi (1985) and then they illustrate this rule with examples from the Siarad corpus. This analysis shows the relevance of using the matrix language concept with the Siarad corpus and it leads to the part focused on the Matrix Language Frame model (Myers-Scotton, 2002) and its three main principles (the asymmetry, the uniform structure, and the matrix language). According to the authors, this model meets the most important set of criteria such as dealing with production data, an ability to analyze the individual clause, and an ability to be used for both bilingual and monolingual clauses. Later in the chapter the authors talk about identifying the matrix language of a given clause with the help of the System Morpheme Principle and the Morpheme Order Principle and applying the MLF model to the data collected from Welsh-English bilinguals. The information is supported by examples from the bilingual corpus and detailed explanations. Also, in chapter 5 the authors provide an overview of the previous studies, where the MLF model was applied to the Welsh-English data (Deuchar 2006, Davies 2010, Deuchar & Davies 2009, Davies & Deuchar 2010), criticism of the MLF (Auer & Muhamedova 2005), and comparisons of the MLF with the Minimalist Program. At the end of the chapter the authors talk about the advantages

and efficiency of the autoglossing system (CLAN) and its efficiency in terms of usage for identifying ML Matrix language within a clause (authors had chosen to use the System Morpheme Principle only) for big sets of data.

Chapter 6 is focused on several independent variables such as age, gender, level of education, etc. and the connection to linguistic behavior. In this chapter, a review of the previous studies dedicated to each of the mentioned variables and the relation between the extralinguistic factors are presented.

Chapter 7 has been divided into three parts, where each part describes a study about change in a certain aspect of Welsh grammar: language shift, convergence, and auxiliary deletion. All these studies are based on the data from the Siarad corpus. In the first study Deuchar & Davies (2009) discuss the future of the Welsh language based on models of grammatical change and code-switching proposed by Alan Thomas, who considered code-switching as an indicator of language decay, and "matrix language turnover" (Carol Myers-Scotton, 1998). In the second study on the Siarad data, Davies & Deuchar (2010) found 26 examples in the corpus and outlined three main types of dichotomous ML clauses, which "show conflicting information as to the source of ML" (Deuchar et al., 2018: 80). The third study turns to the colloquial Welsh, examining the increasing auxiliary deletion (AuxD) in speech. According to Davies & Deuchar (2014), AuxD may have an impact on the future state of Welsh grammar, they also discuss the reasons for AuxD to occur in speech and the factors, influencing its frequency.

Chapter 8 is entitled "Additional research using Siarad", and it touches upon aspects and questions of research where the Siarad corpus already has been used. In this chapter, the reader finds a description of research projects on quantitative and qualitative analyses of code-switching, triggers of code-switching, accommodation in code-switching, and other works connected with spoken Welsh language.

Chapter 9 summarizes the whole book with a detailed description of each chapter. Moreover, the authors discuss possible future directions for the research where the Siarad corpus can be used. Here several ideas on Code-Switching, the Welsh language, the English language, and the Minority languages are suggested.

The appendix section contains the documentation file for the corpus, the list of corpora of the contemporary Welsh before the collection of the data for Siarad corpus, call for participation in the project (Welsh and English versions) and questionnaires for participants as well as consent forms (given in Welsh and English as well). These materials may serve as a model for those who are planning to start their first research project.

To sum up, there clearly exists an increasing interest in spoken corpora focused on code-switching and this book is a good example of the practicality and relevance of contemporary bilingual corpora. The book covers a large variety of topics in

applied linguistics and it may serve as a manual for researchers who are interested in creating or working with a bilingual data corpus with informal conversations. This work is highly recommended for everyone who is interested in bi- or multilingualism from a BA student to a researcher in the field of applied linguistics.

## References

- Davies, P., & Deuchar, M.** (2010) Using the Matrix Language Frame model to identify word order convergence in Welsh-English bilingual speech. In: Breitbarth, A., Lucas, C., Watts, S. & Willis, D. (Eds.), *Continuity and Change in Grammar*. Amsterdam: John Benjamins. 77-96.
- Davies, P., & Deuchar, M.** (2014) Auxiliary deletion in the informal speech of Welsh-English bilinguals: A change in progress. *Lingua*, 143. 224–241.
- Deuchar, M., & Davies, P.** (2009) Code-switching and the future of the Welsh language. *International Journal of the Sociology of Language*, 195. 15–38.
- Deuchar, M., Webb-Davies, P. & Donnelly, K.** (2018) *Building and Using the Siarad Corpus. Bilingual conversations in Welsh and English*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Joshi, A.** (1985) Processing sentences with intrasentential code-switching. In: Dowty, D. R., Karttunen, L. & Zwicky, A. (Eds.), *Natural Language Parsing*. Cambridge: Cambridge University Press. 190–205.
- Myers-Scotton, C.** (1998) A way to dusty death: the matrix language turnover hypothesis. In: Grenoble, L. A. & Whaley, L. J. (Eds.), *Endangered Languages: Language Loss and Community Response*. Cambridge: Cambridge University Press. 289–316.
- Myers-Scotton, C.** (2002) *Contact Linguistics: Bilingual Encounters and Grammatical Outcomes*. Oxford: Oxford University Press.

PENIUSHINA NATALIA  
University of Pannonia  
Multilingualism Doctoral School, PhD student  
natalia.peniushina@yandex.ru

We acknowledge the financial support of Széchenyi 2020 under the **EFOP-3.6.1-16-2016-00015**.