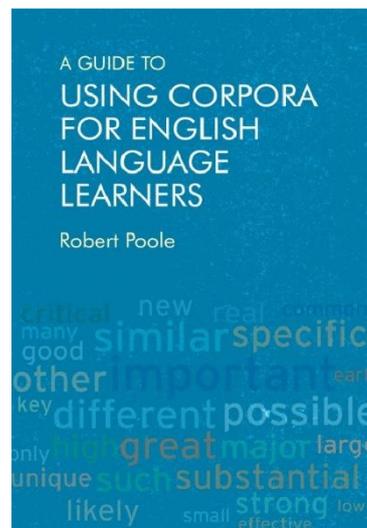Robert Poole

A Guide to Using Corpora for English Language Learners

Edinburgh, Edinburgh UP, 2018, pp.168

Review by Sharon Hartle

Keywords: *freely accessible digital corpora, English language learning, classroom-based activities, concordance searches, grammatical and lexical pattern analysis, socio-cultural language analysis, corpus construction*

A Guide to Using Corpora for English Language Learners is a timely introduction to the world of corpora designed to be a practical starting point for students of English and is written in an accessible style structured by means of a scaffolded step-by-step approach to the topic. The writing is clearly calibrated to be comprehensible to students who have an intermediate level in English and the contents are introduced in a logically staged and structured way. This is a handbook which can be considered to be part of the data-driven learning (DDL) approach, which involves the use of corpora and in particular concordance searches for word pattern analysis to be conducted in class.

Successful language learning is thought to depend crucially on the "activity and initiative of the learner" (Van Lier 2008, 163). Such initiative can be fostered by sensitizing learners to resources such as corpora showing them how to conduct concordance searches. Developing these skills, in fact, encourages independent learning skills, and is thought to foster learner autonomy. The term DDL was coined by Johns (1986; 1991), who advocates an enquiry-based approach where learners' work lies in using corpora themselves to research the language. Analysing lexical and grammatical patterns in this way facilitates student access to the facts of linguistic "performance" (1991, 2) searching for the answers to questions about the way language is used, and the frequent combinations that can be found. This methodology was greeted with initial enthusiasm in language teaching contexts, but over the years became unpopular with some teachers and learners for a range of reasons, such as the difficulty of accessing corpora and the time required to develop the specific skills involved in conducting corpus searches. Another

obstacle lies at times in the student belief that it is their teachers' role to provide answers rather than their own responsibility to explore resources independently. In fact, learner agency is a complex construct which is difficult to define but is often considered as "the socioculturally mediated capacity to act" (Ahearn 2001, 112) but if the sociocultural beliefs of students are not in line with the conviction of their own autonomy conflicts arise.

With the advent of the internet, however, autonomy when studying is becoming the accepted sociocultural norm in many parts of the world. DDL is being revisited, in the light of this, as an effective approach to the study particularly of areas such as collocation and language patterning. In fact, both the advances in technology that have been made over the past few years together with an increased willingness among learners to develop their own independent skills have increased the popularity of DDL. Digital interfaces, such as Sketch Engine's *Sketch Engine for Language Learning* (*SkeLL*) Interface, which facilitate corpora searches, provide the means to conduct simple but effective corpus searches with carefully graded results. These interfaces produce results which often grade language items according to word class and concordance so as to be of immediate benefit to learners. Such interfaces are easily accessible to learners at no cost, factors which are decisive in DDL implementation (Tribble 2015). This volume, which provides a carefully staged introduction to such searches, contributes even further to facilitating the process of using corpora in language learning and taking the mystery out of corpus searches. Teachers and learners, in fact, are increasingly seeing the value of tools like this, that are tailor-made for language pedagogy.

The author, Robert Poole, Assistant Professor of Applied Linguistics and TESOL at the University of Arizona, has written this guide, in fact, to be used either independently by language learners themselves or by educators who seek to include elements of DDL in their own teaching contexts. His introduction to learners explains briefly what corpora are and how they may be useful to those seeking to understand more delicate differences in meaning and usage of lexis. He stresses the fact that this book will foster learner independence gradually. In his introduction to language teachers he addresses issues such as when and how to introduce DDL into teaching moments stressing that "many of the tutorials in this book were developed in response to particular classroom realities such as the need to develop my students' abilities to revise papers or improve word choice" (ix). He also stresses the formative value of giving precise, concrete strategies to learners which will help them to achieve better results rather than simply providing comments such as "Avoid this…" (ix).

The book is structured in five main sections which all develop learner skills in a practical way by providing clear explanations and instructions alternating with "Your Turn" sections, where

learners are encouraged to carry out their own searches based on the models previously provided. The aim, therefore, is to teach the skills of corpus investigation rather than to focus on specific lexical items.

The first chapter focuses on the basics of corpus search functions using *the Corpus of Contemporary American English* (*COCA*), *the British National Corpus* (*BNC*) and *the global Web-based English Corpus* (*GloWbE*) as models. The author points out that whilst use of these corpora is prevalent, there are also many other easily accessible corpora interfaces available online and he provides a comprehensive list of such resources with short descriptions in the section following Chapter 5. The aim of this section, however, is to see what sort of corpus search can be carried out in general, even though the practical steps involved in such searches may vary from interface to interface. This involves the development of an awareness of collocation and language patterns, frequency and comparison between differing usage in different Englishes or genres. The "Your Turn" activities promote autonomy and encourage learners to duplicate the searches to analyze their own choices. The author, however, also advocates the use of corpus searches to understand semantic meaning (8), which is more problematic for learners. Expert users already have a knowledge of the term to draw on but those who are less experienced may find this challenging.

Chapter 2 focuses on corpus searches for vocabulary learning which the author considers to be one of the most common aims of corpus use by learners, particularly when preparing for large scale standardized tests such as TOEFL. The value, he says, of choosing a corpus interface rather than a learner dictionary is that a corpus provides "multiple examples of a word being used in a variety of authentic contexts" (43). In this chapter, he explores several corpora and ways in which they can be used specifically for language learning building on the skills introduced in Chapter 1. One example of this is word formation, which is commonly tested in standardized tests. He illustrates techniques for exploring this area by means of the root '-audi,' related to hear, and encourages learners to do a wildcard search for 'audi*' as a simple 'list search' in *COCA*, producing the most frequent examples from the corpus that contain this root. What might be even more beneficial at this point, however, would be further activities for learners to process and develop their own production of such terms rather than the simple study and understanding of the terms, although that is perhaps beyond the scope of this volume. He also introduces new interfaces in this chapter such as *wordandphrase.info*, which is particularly useful for investigating academic registers, and which enables close reading both of source texts and learners' own texts.

Chapter 3 moves on to focus on the use of corpora to develop more effective writing skills. The activities in this chapter are based directly on ones that the author has used in his own classes and are, therefore, relevant to a specific group of learners who need to write in an academic register, although the activities can be adapted for writing in other contexts. The activities cover a range of aims such as helping to identify error, improving word choice to working on the sub-skills of writing. These include the use of conjunctions for coordination, which are directly related to language learning strategies and work on determining the register of phrasal verbs, which generally prove to be uncommon in academic contexts, for example. The author suggests comparing the text types where a phrasal verb such as 'figure out' may naturally be found, with the contexts where its synonym 'determine' is found. The phrasal verb occurs mainly in the spoken language, in fiction or in magazines, all of which tend towards a more informal register. This is not the case with 'determine' which tends to be more formal or academic. To do this, once again, entails learners' prior awareness of those synonyms, which may not be the case and cannot be taken for granted. Therefore, it would be useful to add awareness-raising activities at this stage, as well.

Chapter 3 also contains activities to help learners identify example articles on topics that are being written about and activities aimed to enable learners to observe variation of register and writing practices in different disciplines. The main corpora referred to in this section are *the News on the Web* (NOW) *Corpus* as well as *COCA, the Michigan Corpus of Upper-Level Student Papers* (MICUSP), *wordandphrase.info*, SkeLL and *the corpus-based FLAX Interactive Language Learning site.*

Chapter 4 initially moves away from purely language learning concerns to explore resources that may be used to prompt and foster critical thinking. The activities go beyond the items being explored to see what usage reveals about its users from a sociocultural viewpoint. This is directed at learners but also at teachers, who may wish to develop such searches to inform their own teaching practice. The author suggests using searches that reveal sociocultural information as a basis for topics to be explored in conversation classes, for instance. One example of this is the use of *the Google Ngram Viewer* to see the chronological trends in lexical use, which he introduces with an activity that explores the use of 'teenagers,' 'zombies' and 'selfies' from 1900 to 2000 (121). Learners are asked to analyze the results and then discuss the reasons they think may lie behind those results and the author also introduces *the Corpus of Historical American English* (*COHA*) to aid such discussion.

The chapter, however, later returns to language learning matters by focusing on the identification of language use patterns by adopting *COCA* collocate searches. This enables

learners to see how nouns are generally modified and which adjectives, for instance, generally modify 'man' and which ones generally modify 'woman.' This may lead to both discussion and conversation in class as to language issues such as which collocations are considered to be natural when writing or speaking.

The final chapter, Chapter 5, directed both at learners and teachers, deals with building personalized corpora. To do this Poole introduces *AntConc,* the freely available desktop application created and developed by Laurence Anthony which is widely used both because of its ease of access and simplicity of use. Creating a personalized corpus can provide valuable insights into specific types of writing. The author mentions letters to shareholders in Fortune 500 companies, for instance, which could provide useful language information for those who need to write such letters. The chapter deals, in a very practical way, with how to organize texts to be included in a corpus, how to 'tidy' them and import them into *AntConc* and how to conduct searches once the corpus has been created. This section is followed, as was mentioned above, by a useful list of freely available corpus resources for teachers and learners. Further reading resources are also provided, although there are only four texts listed and they may be useful for teachers but are probably not practical enough for use by learners. The book is well-designed and its practical approach will be useful both to educators and learners. The step-by step development provides a clear, comprehensive path for students, introducing them to this world by means of guided discovery and showing them how they can improve their own language skills by using easily accessible digital interfaces to search the different corpora. The introduction would, however, benefit from clearer referencing of bibliographical sources, since the volume is directed at both learners and educators and the latter may well wish to read more about the topic. At times, as has been mentioned above, certain steps could be added to make the searches even more useful to learners, particularly when dealing with semantic meaning or synonyms. It would be useful to explore other tools which can be used together with the corpus searches. Such work, however, may be beyond the scope of the volume and the activities in the book are, in any case, an extremely valuable resource.

Sharon Hartle *is an associate professor in the Department of Foreign Languages and Literatures at Verona University. She is specialized in English Language teaching (ELT) pedagogy and didactics and works specifically in the field of English for Specific Purposes (ESP). She has worked for some years in the field of e-learning and specializes in multimedia material development for ELT in Blended Learning contexts. Her research interests also extend to include English Language Assessment and English Medium Instruction (EMI).*

Works cited

Ahearn, Laura. "Language and Agency." *Annual Review of Anthropology* 30 (2001): 109–137.

Johns, Tim. "Micro-Concord: A Language-Learner's Research Tool." *System* 14.2 (1986): 151-162.

---. "Should You Be Persuaded: Two Samples of Data-Driven Learning Materials." *English Language Research Journal* 4 (1991): 1-16.

Tribble, Christopher. "Teaching and Language Corpora Perspectives from a Personal Journey." *Multiple Affordances of Language Corpora for Data-driven Learning.* Edited by Agnieszka Leńko-Szymańska and Alex Boulton. Amsterdam: John Benjamins, 2015. 37-64.

Van Lier, Leo. "Agency in the Classroom." *Sociocultural Theory and the Teaching of Second Languages.* Edited by James P. Lantolf and Matthew E. Poehner. London: Equinox, 2008. 163– 86.

Corpus Tool Interfaces cited

AntConc, http://www.laurenceanthony.net/software.html.

FLAX, *http://flax.nzdl.org/greenstone3/flax?a=fp&sa=collAbout&c=collocations.*

The British National Corpus (BNC), https://www.english-corpora.org/bnc/.

The Corpus of Contemporary American English (COCA), https://www.english-corpora.org/coca/.

The Corpus of Historical American English (COHA), https://www.english-corpora.org/coha/.

The Global Web-based English Corpus (GloWbE), https://www.english-corpora.org/glowbe/.

The Google Ngram Viewer, https://books.google.com/ngrams.

The Michegan Corpus of Upper-Level Student Papers (MICUSP), http://micusp.elicorpora.info/.

The News on the Web Corpus (NOW), https://www.english-corpora.org/now/.

Sketch Engine's Sketch Engine for Language Learning (SkeLL), https://skell.sketchengine.eu/#home?lang=en.

Wordandphrase, https://www.wordandphrase.info/.