Retrospectiva

MCCARTEN, J. **Teaching Vocabulary**: lessons from the corpus, lessons for the classroom. Cambridge: Cambridge University Press, 2007.

Raphael Marco Oliveira Carneiro[*]

Teaching Vocabulary: Lessons from the Corpus, Lessons for the Classroom, written by Jeanne McCarten and published by Cambridge University Press is a booklet of no more than 30 pages. Notwithstanding its few pages, it is a work of paramount importance on what concerns the use of corpus in language teaching. That is why this book is worth reading. Not only does it address fundamentals of corpus linguistics and how information about language derived from corpus research can be used to inform vocabulary teaching, but it also suggests practical classroom approaches to foster vocabulary learning and comprehension. Besides this, the content is organized in a simple and direct manner. There are four main chapters: *1. Lessons from the Corpus, 2. Lessons for the Classroom, 3. Concluding remarks and 4. Appendices*. I shall then make a brief description of each chapter.

The first chapter is the biggest one and is split in nine sections. It addresses some basic concepts on corpus and vocabulary issues. Right at the outset, the following two questions are posed: "How many words are there and how many words do we need to teach?" and "What can a corpus tell us about vocabulary?"

The author attempts to answer the first question raising important concepts on vocabulary learning, such as the number of words one needs to know in order to get along in English. According to some research "learners who know the most frequent 2,000 words should be able to understand almost 80 percent of the words in an average text, and a knowledge of 5,000 words increases learners' understanding to 88.7 percent" (MCCARTEN, 2007, p. 1). On what concerns spoken language, according to the author, an understanding of 1,800 words would be enough as word repetition is much higher in speech. From these considerations, it is possible to point out the first lesson derived from a corpus: to identify and give priority to the most frequent 2,000 and 5,000 vocabulary items in teaching.

---

[*] Undergraduate student of the English Language and Literature Course of the Federal University of Uberlandia; e-mail: raphael.olic@gmail.com.

The second question is answered by taking into account the concept of corpus. According to the author:

> A corpus is basically a collection of texts which is stored in a computer. The texts can be written or spoken language. Written texts like newspapers and magazines can be entered into the computer from a scanner, a CD, or the Internet. Spoken texts, like conversations, are recorded and then the recordings are transcribed; that is, they are written down word for word, so that the texts of these conversations can be fed into the computer database. It is then possible to analyze the language in the corpus with corpus software tools to see how people really speak or write (MCCARTEN, 2007, p. 2).

Bearing this concept in mind, as well as the one that a corpus has to be designed according to principled criteria depending on the aims of its creation, we can say that a corpus can basically tell us the following information about vocabulary:

> **Frequency:** Which words and expressions are most frequent and which are rare
> **Differences in speaking and writing:** Which vocabulary is more often spoken and which is more often written
> **Contexts of use:** The situations in which people use certain vocabulary
> **Collocation:** Which words are often used together
> **Grammatical patterns:** How words and grammar combine to form patterns
> **Strategic use of vocabulary:** Which words and expressions are used to organize and manage discourse (op. cit., p.3).

Thus, a corpus can provide us data for both quantitative (statistical information, frequencies) and qualitative analyses (observe how people use vocabulary in context). In order to carry out these analyses, it is necessary to use corpus tools such as the lexical analysis software WordSmith Tools[1] (SCOTT, 2008).

McCarten further discusses each aspect mentioned above illustrating each one of them with examples extracted from conversations found in the North American spoken corpus, which is part of the *Cambridge International Corpus*. The examples of corpus-informed activities are taken from the textbook series *Touchstone*.

Amongst all the aspects aforementioned, it is important to highlight the one of strategic vocabulary that as far as I am concerned is underexplored in language teaching. Furthermore, it is pointed out the importance of finding a vocabulary of conversation that reflects the interactive nature of conversation and that gives it its distinctive character. Language features such as discourse markers, responses, monitoring expressions, vague expressions, hedging

---

[1] The WordSmith Tools, developed by Mike Scott, is currently in its 6th edition and it can be downloaded from the following link: <www.lexically.net>.

expressions and expressions of stance are not usually given the importance they deserve in the classroom. It is beyond the scope of this review to characterize and exemplify each one of them. Suffice it to say that the practice of this kind of vocabulary can make students not only sound more natural and fluent, but also able to manage and conduct conversations.

Still in the first chapter, it is discussed how strategic vocabulary can be fitted into materials and the course syllabus. According to the author, a syllabus that intends to cover issues on conversation strategies and skills ought to include the following functional areas: organizing your own talk, taking account of another speaker, showing listenership, that is showing you understand by responding appropriately, and managing the conversation as a whole. As McCarten remarks, "Mastery of these four aspects of conversation helps speakers, and therefore learners, to participate in and manage successful, fluent conversations" (MCCARTEN, 2007, p. 14).

So much for the first chapter, the second one discusses a wide range of possibilities to foster and boost vocabulary learning. The discussion is mainly guided by two questions: "What do we need to teach about vocabulary?" and "How can we help learners learn vocabulary?"

The author, then lists the following aspects that learners need to know about a word: the meaning(s) of the word, its spoken and written forms, what "word parts" it has (e.g., any prefix, suffix, and "root" form), its grammatical behaviour (e.g., its word class, typical grammatical patterns it occurs in), its collocations, its register, what associations it has (e.g., words that are similar or opposite in meaning), what connotations it has, and its frequency. By the same token, teachers must consider the kind of vocabulary students will need to learn, be it for either active or productive purposes, or passive or receptive purposes.

The second question is addressed by means of suggesting principles that teachers might use in order to help students learn vocabulary more effectively, such as: focus on vocabulary, offer variety, repeat and recycle, provide opportunities to organize vocabulary, make vocabulary learning personal, do not overload students with too much vocabulary and use strategic vocabulary in class. Likewise, the author points out the importance of helping students become more independent learners in and out of class by using strategies such as vocabulary notebooks and research tools. For instance, learners' and online dictionaries, as well as the Internet are invaluable sources of information that can be exploited in order to learn more about collocations, grammatical patterns, idiomatic expressions and the like. Learners can even access

corpora available online such as the *Corpus of Contemporary American English*[2] (COCA), in order to search examples of language in use.

Ultimately, by the end of the book it is acknowledged that the acquisition of vocabulary is arguably the most critical component of successful language learning, as well as that corpus research has given access to a wealth of information on spoken and written English that was previously unavailable. Thus, owing to this fact, teachers are posed the challenge of finding ways to make use in the classroom of the new information provided by corpora.

In the appendices, it is included a list of the top 200 spoken words, suggestions for further reading on both corpus and vocabulary issues, as well as the references used throughout the booklet. Having described the main points of the book, I shall make some comments considering the book as a whole.

First and foremost, the usefulness of this book resides in the new insights provided on what concerns vocabulary teaching, chiefly on the inclusion of strategic vocabulary use. It is not aimed at being a comprehensive book on the subject, but the new concepts discussed in it are likely to help teachers reflect upon what kind of vocabulary they teach and how they teach vocabulary in the classroom. It is clear that the key issue is how to fit the information provided by corpus research in the language classroom. And in this way, the book serves its purpose by giving some directions on how information derived from corpus can be incorporated in vocabulary teaching. Therefore, it is significant that teachers start considering the possibility and the availability of corpus use in the classroom to enhance the learning experience.

Moreover, this booklet is likely to be of interest mainly to English teachers, graduate and undergraduate students alike who are not familiar with the use of corpus in teaching and want a simple and brief introduction on corpus linguistics and vocabulary teaching.

All things considered, a corpus can teach us the kind of vocabulary people use in both written and spoken texts on a wide range of topics. Over and above that, language can be investigated through the use of corpus in many different aspects. Thus, although a corpus can be a very useful tool, it should not be regarded as the be-all and end-all of teaching; nor should it be deemed as a panacea for the learners' linguistic needs. In the words of McCarten (2007, p. 3):

---

[2] The COCA was created by Mark Davies of Brigham Young University and it is currently composed of 450 million words from 1990 to 2012. It is available on the following link <http://corpus.byu.edu/coca/> (accessed on Sep. 20, 2013).

> [a corpus] cannot tell us exactly what to teach or how to teach, and it has nothing to tell us with respect to how students learn best. It cannot replace the expertise of teachers, or of students themselves, on how best to teach and learn vocabulary. It is *a* tool. It is not the *only* tool.

As the above extract bespeaks, the underlying notion of all the issues discussed by the author is that owing to the myriad of words of the English language, teachers are unlikely to cover them all in class. Thereby, a corpus can help them identify the most important words to be taught; as well as provide some important information to help learners become more fluent. Furthermore, it is due to the huge enterprise of learning vocabulary that it is equally important to help students not only with what to learn, but also with how to learn, so that they can become more autonomous learners and more capable of ensuring the continuity of their learning outside the class.

References

MCCARTEN, J. **Teaching Vocabulary**: Lessons from the Corpus, Lessons for the Classroom. Cambridge: Cambridge University Press, 2007.

SCOTT, M. **WordSmith Tools**. Version 5. Liverpool: Lexical Analysis Software, 2008.