

## AUTHOR/EDITOR QUERIES

ARTICLE ID : LLC-FQQ010			
Query No	Section	Paragraph	Query
Q1	Author names		Please check that the author surnames (family name) have been correctly identified by a pink background. If this is incorrect, please identify the full surname of the relevant authors. Occasionally, the distinction between surnames and forenames can be ambiguous, and this is to ensure that the authors' full surnames and forenames are tagged correctly, for accurate indexing online.



# Review

---

Exploring English with Online Corpora.  
An Introduction.  
5 Wendy Anderson and John Corbett.  
*Basingstoke: Palgrave MacMillan, 2009. 224 pp.*  
*ISBN 9780230551398. £49.50 (hardback).*

10 This textbook by Wendy Anderson and John  
Corbett is presented as a low threshold introduction  
into the main linguistic properties of English based  
on the use of online corpora. It addresses a  
non-specialist readership. The book is subdivided  
15 into eight chapters that successively deal with the  
nature and analysis of English corpus data  
(Chapters 1 and 2); lexical, grammatical, discourse,  
and phonetic properties of English that can be  
explored through online corpora (Chapters 3, 4, 5,  
20 and 6), the contextualization of corpus data and  
their use in education and research (Chapters 7  
and 8). There is an appendix with a brief description  
of 18 online corpora, a term list explaining the main  
technical terms used in the book, a reference list,  
25 and an index. Each chapter clearly defines its objec-  
tives and identifies a number of interesting research  
questions, which are then answered on the basis of  
results yielded by corpus searches. Readers are also  
set a number of tasks, which are mainly searches  
30 they have to carry out themselves to reach a conclu-  
sion about the use of particular linguistic items and  
structures. Each chapter suggests further readings  
on the topics it discusses.

35 The book covers the main areas of linguistic  
research in an order that generally reflects the  
order of introductory courses of linguistics, starting  
with small linguistic units and building up to higher  
levels of analysis, involving discourse and social  
parameters in speech and writing. However, this  
logical structure is breached by the chapter on pro-  
40 nunciation (Chapter 6), which is located between  
the chapters on discourse and context properties.

This somewhat unusual position is probably due  
to the fact that lexical, grammatical, and discourse  
properties can be investigated in all available cor-  
pora, whereas phonetic information requires special  
corpora of spoken data.

45 The first two chapters provide a good general  
introduction to corpora, the history of corpora in  
linguistic research, the analysis of corpus data, and  
some basic statistical tools to interpret search  
50 results. The authors insist on the main issues of  
corpus design, such as types of corpora and repre-  
sentativeness and stress the importance of both  
quantitative and qualitative data. In a research  
field that is traditionally dominated by numbers  
55 and frequencies, it is indeed important to remind  
that a qualitative approach can usefully supplement  
the quantitative approach. The authors also quite  
rightly insist on the limitations of corpus-based  
research and the importance of a critical mindset.  
60 Though the first chapter briefly discusses the history  
of corpora in linguistics, no reference is made to the  
important role played by corpus linguists in the  
discipline and the scientific achievements made on  
the basis of corpora. The main corpora that are used  
65 throughout the book are also presented in the  
first chapter: the British National Corpus (BNC)  
which is available through two different websites,  
the Corpus of Contemporary American English,  
the TIME Corpus, the Michigan Corpus of  
70 Academic English, and the Scottish Corpus of Text  
and Speech (SCOTS). The latter of these corpora  
was compiled by the authors of the book, which  
probably explains why it is so prominent in the  
following chapters.

75 Chapters 3 and 4 deal with the corpus-based  
investigation of specific low-level linguistic items  
and structures. Chapter 3 on lexical items discusses  
basic questions in lexicography such as the status of  
a word in a corpus, its relation to a lemma and its

Review

involvement in collocations, colligation, and semantic prosody. Chapter 4 on grammar provides an overview of grammatical concepts, such as parts of speech and grammatical functions. The focus in these two chapters is on the use of corpora to find information on specific combinations of words or categories and not on the theoretical problems that arise in relation with these concepts. In a linguist's view this looks like a missed opportunity, especially if working with corpora. The grammatical categories used in the book are in fact no more refined than those used in most school grammars. However, one of the achievements of corpus-based linguistics is precisely to have shown that these categories are problematic. The authors do occasionally allude to problems of that kind (see for instance their comment on adjectives used as adverbs on p. 76), but no systematic attempt is made to make the readership aware of the intuitive nature of traditional grammatical categories. This approach also reflects in the many hedged statements in Chapter 4 ('usually', 'normally', and 'generally') which look strangely imprecise for a book that speaks highly of corpora as a means of obtaining quantified information on language.

Chapter 5 on discourse draws attention to one of the shortcomings of many online accessible corpora, i.e. the fact that only short fragments of texts can be accessed through search items. As discourse properties are properties of large portions of texts, their systematic study is hindered by the format in which search results are shown. To study discourse properties, one therefore needs a corpus from which entire texts can be downloaded. The authors then focus on items such as adjacency pairs and cohesive devices (pronouns and conjunctions).

Chapter 6 discusses several aspects of English pronunciation, including phonetics and prosody and, more in particular, to the differences corpora allow us to detect between distinct varieties of English. The authors provide a basic introduction into phonetics (IPA) and phonology (wrongly defined as the study of the combination of sounds in patterns, p. 125) and invite readers to check the pronunciation of vowels and consonants in different

varieties of English, especially the Scottish variety they have recorded in the corpus they compiled.

Chapter 7 stresses the importance of metadata and their use in corpus-based research on variation. It distinguishes between user-related data and use-related parameters and illustrates how these parameters can determine linguistic choices which are reflected in the corpus data. In fact, several of these parameters also appear in other chapters. Chapter 7 thus functions as a kind of round-up of the preceding chapters

Finally, Chapter 8 suggests some ways of working with corpora in the classroom and in research and looks out into the future of corpora. Unfortunately, the suggestions do not add much to what was already illustrated in the previous chapters and the part on research does little more than referring to the empiricist-generative divide in linguistic research. It would have been more interesting to illustrate how particular research questions asked by generative linguists can be tackled with corpus research.

My general impression of the book is that it certainly is an excellent introduction into corpus research for readers without prior knowledge of linguistics. It provides a wealth of exercises and intelligently invites the reader to further reflect on the results of the searches. For readers with some knowledge of linguistics, many of the proposed topics will probably be too anecdotic. The book also successfully puts the corpus compiled by both authors (SCOTS) in the spotlight. The corpus no doubt deserves attention, as it contains a wealth of information, including a feature no other online corpus provides: clickable maps indicating where occurrences of particular items were recorded. Linguists might therefore be interested in using the book as an introduction to the corpus.

Bart Defrancq  
Translation Studies,  
University College Ghent,  
Ghent University,  
Belgium  
doi:10.1093/llc/fqq010

Q1