Open Research Online

Review of: *'New Trends in Corpora and Language Learning' and 'Keyness in Texts'*

Journal Item

For guidance on citations see FAQs.

oro.open.ac.uk

Frankenberg-Garcia; Flowerdew, Lynne; & Aston, Guy (Eds) 2011 New Trends in Corpora and Language Learning. Continuum, London and New York. 268 pp.

Bondi, Marina & Scott, Mike (Eds) 2010 Keyness in Texts. John Benjamins, Amsterdam and Philadelphia. 251 pp.

New Trends in Corpora and Language Learning provides exactly what you would expect from the title: a comprehensive update on current issues within the rapidly-growing field of corpus use in language learning and teaching. Originating from research first reported at the 2008 Teaching and Language Corpora (TaLC) conference held in Lisbon, the 15 chapters are shared across three sections: 'corpora *with* language learners: use', 'corpora *for* language learners: tools', and 'corpora *by* language learners: learner language' (my emphasis). The diversity of the studies in the book presents something for every reader – whether you are an experienced language teacher but novice user of corpora, a corpus linguist who is unfamiliar with language teaching, or perhaps a corpus software developer who wishes to learn about current concerns in language learning.

The first of the three sections, 'corpora *with* language learners: use', offers practical ideas and empirical studies on how to make use of corpora in the classroom with language learners who have no previous experience in exploring large quantities of text using a computer and for whom concordance lines, collocate analysis and frequency lists may be alien concepts. A couple of chapters stood out for me. In Chapter 1, Yukio Tono kicks off with a discussion of corpora use in the Japanese context and introduces us to the quirky cartoon character 'Mr Corpus'. Disconcertingly resembling a wooden spoon in the accompanying picture (p.9), 'Mr Corpus' appears in an English TV series in Japan named '100-go'. Each programme features one of the most frequent 100 words in English language vocabulary (the basis for this frequency list is the spoken part of the BNC). Occasional appearances are also made in the TV programmes by 'Dr Corpus' (a.k.a. Yukio Tono himself). More than three million Japanese people have watched '100-Go', turning 'corpus' into a 'buzz word' (p.10) and helping corpus-based English language material to gain popularity in Japan.

In contrast with the nation-wide, high-profile 'Mr Corpus' in Japan, Maggie Charles (Chapter 2) gives us a small-scale account of corpus use in one teaching context in the UK. She provides a highly practical account of how to use concordance lines in an English for Academic Purposes classroom with mixed-discipline, non-linguist groups of university students. Her approach is very accessible, and should give any teacher who is new to corpus linguistics the confidence to try it out. Charles presents ideas as to the kinds of questions to put to the data (e.g. how do writers show criticality?), the number of concordance lines to use (10-20 lines, expandable as more context allows patterns to become clear), and also puts forward a convincing case for issuing students with paper concordances of the data (as a record). Following a questionnaire study of her students' views and a discussion of using corpora within mixed discipline classes, Charles concludes that '[w]hat students learn from their concordancing is the [...] skill of noticing and interpreting linguistic features and this can be profitably applied in their own fields' (p.37). Students learn that the corpus does not have the 'right' answer – and that language may work a little differently in their (sub)

discipline and within a particular genre - but also that an academic corpus is a valuable source of data for the student as language analyst to grapple with and learn from.

The rest of section one looks at corpus analysis of 'Emo' texts within cultural studies (Kettemann, Chapter 3), corpora used by trainee translators (Kübler, Chapter 4), and a free online concordancing tool used with Polish EAP students (Kaszubski, Chapter 5). From these examples of corpus use with language learners, we move to section two in which the focus is the use of 'tools' for exploring corpora. Here we are introduced to a tool for exploring verb-noun miscollocations ('Collocation Explorer', Liu, Wible and Tsao, Chapter 6); an English-Spanish Content-Based Machine Translation project (Fuentes, Chapter 8); and 'ConcGram' (2009, Greaves), a tool for identifying phraseologies (discussed by Warren, Chapter 9). Nestled between these studies is Francesca Coccetta's (Chapter 7) consideration of multimodal analysis focusing on online spoken corpora. Coccetta discusses the 'common practice of applying approaches borrowed from the investigation of corpora of *written* texts' (p.121 original emphasis), arguing convincingly for the inclusion of facial expressions, gestures, tone of voice and so on in studies of spoken language to enable us 'to see how language interacts with the other semiotic resources' (p.126). The importance of this 'multi-modal co-text' (Baldry, 2008) when examining concordance lines is stressed throughout the chapter within an approach focusing on notions and functions in text. The final chapter in section two should be sufficient on its own merit to convince secondary school foreign language teachers of the usefulness of corpora in the classroom. Here, Widmann, Kohn and Ziai (Chapter 10) propose a range of ideas for classroom activities. Framed within a study of video interviews of secondary school students, they illustrate how teenagers can be motivated to learn through use of a corpus tool which allows searches for topics, word patterns or language variety, and enables the user to watch video interaction whilst reading annotated transcripts.

The final third of the book deals with 'learner language'. This is fairly loosely defined since it includes the successful undergraduate student assignments within the British Academic Written English (BAWE) corpus even though these have all achieved a IIi or I (over 60%) in the UK system and represent proficient writing at this level by both native and non-native speakers (Nesi, Chapter 13). Also in the third section, Osborne (Chapter 11) examines oral learner corpora, asking if there is such a thing as 'overall fluency' and, assuming this exists, how it can be measured. Both quantitative and qualitative measures are considered with pausing and speech rate among the features discussed within the former, and syntactic density, accuracy and lexical range among the latter. Osborne argues convincingly that no single measure of fluency equates to a high level of proficiency (using the Common European Framework of Reference), and posits instead that a bundle of fluency measures is required. Unlike many studies of learner corpora, Osborne's approach does not frame L2 learners' language within a deficit model, but instead points out the diversity of fluency patterns. Notably he states that '[t]here is no measure for which there is no overlap between native and non-native speakers (i.e. there is always a non-native speaker who performs 'better' than at least one of the native speakers of the same language)' (p.189).

Other studies of learner language in this section focus on positive and negative evaluative adjectives (e.g. *good, great, nice, awful, terrible*) (De Cock, Chapter 12), persuasive devices (e.g. boosters such as *of course,* hedges: *possibly,* and attitude

markers: *hopefully*) (Hatzitheodorou and Mattheoudakis, Chapter 14), and discuss revisions made to Portuguese academics' texts before publication within an international journal (McKenny and Bennett, Chapter 15).

Taken as a whole, *New Trends in Corpora and Language Learning* not only brings the corpus linguist – whether novice or expert - firmly up to date with current developments in corpus use within language learning and teaching, it also inspires both researchers and practitioners (and most readers probably possess elements of the two) to act on the wealth of ideas it contains. One useful addition – and an area I always feel is lacking in edited collections of papers using corpora – would be a guide or grid listing the corpora, corpus tools, language and student level employed in each study. Such a guide would offer the reader an alternative pathway through the book, and particularly appeal to anyone with a specific agenda in mind.

While *New Trends* concentrates on corpora in language learning and teaching, the second book considered in this review takes the popular corpus linguistic technique of keyword analysis as its focus. A 'keyword' is 'an ordinary word which happens to be key in a particular text' (Scott and Tribble, 2006: 78); keyness is established by the frequency of a word (or 'cluster' comprising two or more contiguous words) in one corpus with the frequency of the same item in a larger 'reference corpus'. Scott and Tribble (2006: 56) memorably describe keyness as '[w]hat the text 'boils down to' … once we have steamed off the verbiage, the adornment, the blah blah blah', though this assumes a text-based notion of relevance when perhaps the broader environment of the discourse the texts appear in should also be considered. The first few chapters in *Keyness in Texts* expand further on the nature of keyness and what it can reveal.

Keyword analysis (henceforth: KWA) can indicate 'positive' and also 'negative keywords' (items appearing *less* frequently than would be expected by chance). Scott (p.43) refers to keyness as 'a new territory…[which] has attracted colonists and prospectors'. A keyword is viewed as an 'enabling device' (Scott p.44) which helps the researcher to find out what the text is about ('aboutness') or how it is written (style). The notion of keywords is related to Williams' (1976/1983) concept of key words being culturally important content words, though in corpus linguistics the set of candidate keywords is broadened to include all words purely on grounds of greater than expected frequency levels.

Nicholas Groom gives three positives for investigating corpora using KWA: first, this approach avoids researcher bias as a computer simply does as it is asked rather than preferentially selecting particular items for more in-depth analysis; second, KWA is 'unencumbered by previous linguistic theories or descriptions' held by the researcher (p. 60), and third, the procedure uncovers patterns which could not have been predicted in advance (cf. Scott and Tribble, 2006). Keyword analysis thus encapsulates the corpus-driven approach (Tognini-Bonelli, 2001), since keywords merely 'point at elements that need to be explained' (Bondi, p.14). Groom questions the usual divide of keywords into open class items revealing the 'aboutness' of a text and closed class items indicating its 'style', asserting that this is 'missing the point' (p.63) since a focus on closed class keywords reveals a great deal about the content of a text and allows the analysis of Hunston's (2008) semantic sequences.

The remainder of *Keyness in Texts* comprises a set of empirical studies exploring keyness in various contexts, illustrating the range of applications for KWA. Thus we encounter keywords in political speeches (Milizia), travel writing (Gerbig), environmental issues in media coverage of the Kyoto Protocol (Bassi), metaphor analysis (Philip), school textbooks and the 'we-ness' of the national consciousness in Korea (Fraysse-Kim). By far the most popular software used in the studies is Scott's WordSmith Tools, though the ConcGram tool (2009, Greaves) is introduced by Martin Warren (incidentally the only writer to appear in both volumes in this review).

One point I noticed in reading through this volume is that, despite its narrow focus, *Keyness in Texts* contains little discussion of how or why keyness is calculated using the statistical tests of log likelihood or chi square. While these tests establish levels of confidence in the frequency of items in a study corpus compared to a reference corpus, they do not reflect the size of the frequency difference and concentrate only on its statistical significance (cf. Kilgarriff, 2009, see also Gabrielatos and Marchi, 2011). This is a recent area of discussion within KWA and will no doubt be explored in more depth in future papers.

Taken together, these two books add much to the exciting field of corpus linguistics. Whether you are a teacher, linguistics student, second language learning researcher, software developer or simply interested in keeping up to date in an important and fast-changing area of linguistics, each book contains a range of topics of interest. While *New Trends* could be read by the non-corpus linguist and would hopefully inspire them to find out more about corpora in language learning, *Keyness in Texts* is aimed at researchers with at least some experience of conducting or reading research within corpus linguistics. Although no previous knowledge of the keyword procedure is required, *Keyness in Texts* is a specialised volume and the corpus virgin might struggle to follow the detail in these studies. The range of areas of research across the two books in terms of students, languages, applications, tools, disciplines and genres illustrates the array of studies in the field. Yet, despite this diverse output of corpus linguistic research, we are still only scratching the surface of the possibilities available to future researchers in the field. As Scott (*Keyness in Texts*, p.56) succinctly puts it: '[t]here is so much more to do'.

Baldry, A. P. (2008). Turning to multimodal corpus research for answers in a language-course management crisis. In C. Taylor Torsello, K. Ackerley & E. Castello (Eds.), *Corpora for University Language Teachers* (pp. 226-237). Bern: Peter Lang.

Burnard, L. (2009). Reference Guide for the British National Corpus. XML. Retrieved 18/11/2011, from http://www.natcorp.ox.ac.uk/corpus/

Gabrielatos, C., & Marchi, A. (2011). *Keyness: Matching metrics to definitions*. Paper presented at the Corpus Linguistics in the South: Theoretical-methodological challenges in corpus approaches to discourse studies - and some ways of addressing them. Retrieved from http://eprints.lancs.ac.uk/51449/4/Gabrielatos_Marchi_Keyness.pdf.

Greaves, C. (2009). ConcGram 1.0: A phraseological search engine. Amsterdam: John Benjamins.

Hunston, S. (2008). Starting with the small words: Patterns, lexis and semantic sequences. *International Journal of Corpus Linguistics, 13*(3), 271-295.

Kilgarriff, A. (2009). Simple maths for keywords. In M. Mahlberg, V. González-Díaz and C. Smith (Eds.), *Proceedings of the Corpus Linguistics Conference*. Retrieved from http://ucrel.lancs.ac.uk/publications/cl2009/#papers.

Scott, M., & Tribble, C. (2006). *Textual Patterns: Key Words and Corpus Analysis in Language Education*. Amsterdam: John Benjamins.

Scott, M. (2011). *WordSmith Tools, 5.0* [Computer Software]. Oxford: Oxford University Press.

Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work.* Amsterdam: John Benjamins.

Williams, R. (1976/1983). *Keywords: A Vocabulary of Culture and Society*. London: Fontana.

Maria Leedham
m.e.leedham@open.ac.uk
Lecturer in English Language Studies and Applied Linguistics
Faculty of Education and Language Studies
The Open University
Level 2, Stuart Hall
Walton Hall
Milton Keynes
MK7 6AA