

# Corpora and language education

Zhonghua Xiao, Tony McEnery  
(Department of Linguistics and MEL, Lancaster University, UK)

**Abstract:** Corpus linguistics as a methodology of linguistics research has gained such prominence over time that corpora have been used extensively in nearly all branches of linguistics. This article explores, through a series of studies undertaken so far, the potential uses of corpus data in one of these areas – language education. We will discuss a wide range of issues related to using corpora in language pedagogy, including syllabus design, materials development, data-driven learner (DLL), teaching language for specific purposes, language testing, teacher training as well as learner corpus and interlanguage analysis.

**Key words:** corpora, language pedagogy, data-driven learning, learner corpus, interlanguage analysis

## 1. Introduction

The corpus-based approach to linguistics and language education has gained prominence over the past four decades, particularly since the mid-1980s. This is because corpus analysis can be illuminating ‘in virtually all branches of linguistics or language learning’ (Leech 1997: 9; cf. also Biber et al 1998: 11). One of the strengths of corpus data lies in its empirical nature, which pools together the intuitions of a great number of speakers and makes linguistic analysis more objective (McEnery and Wilson 2001: 103). Nowadays, corpora have had such a strong influence on recent reference publishing (at least for English, e.g. Sinclair et al 1990; Francis et al 1996, 1998; Biber et al 1999) that ‘even people who have never heard of a corpus are using the product of corpus-based investigation’ (Hunston 2002: 96). Hunston summarizes the changes brought about by corpora to dictionaries and other reference books in terms of five ‘emphases’: an emphasis on frequency, an emphasis on collocation and phraseology, an emphasis on variation, an emphasis on lexis in grammar and an emphasis on authenticity (*ibid*). In addition to reference publishing, corpora have been used extensively in nearly all branches of linguistics including, for example, lexicographic and lexical studies, grammatical studies, language variation studies, contrastive and translation studies, diachronic studies, semantics, pragmatics, stylistics, sociolinguistics, discourse analysis, forensic linguistics, and language pedagogy. Corpora have won widespread popularity over time in spite of the fact that they still occasionally attract hostile criticism, for example, from scholars such as Widdowson (1990, 2000). In this article, we will not be concerned with the debate over the use of corpus data in linguistic analysis and language education. In our view, such a debate is over a non-issue. Readers interested in the pros and cons of using corpus data should refer to Sinclair (1991), Widdowson (1991, 2000), de Beaugrande (2001) and Stubbs (2001).<sup>1</sup> Nor will we discuss the use of corpora in a wide range of language studies. Readers can refer to Hunston (2002) and McEnery et al (2005) for a

further discussion of using corpora in applied linguistics. This article focuses only on using corpora in language education.<sup>2</sup>

[This is where CBLs starts, p.97 :]

The early 1990s saw an increasing interest in applying the findings of corpus-based research to language pedagogy. The upsurge of interest is evidenced by the six well-received biennial international conferences on *Teaching and Language Corpora* (TaLC).<sup>3</sup> This is also apparent when one looks at the published literature. In addition to a large number of journal articles, at least nine single-authored or edited volumes have recently been produced on the topic of teaching and language corpora: Wichmann et al (1997), Kettemann and Marko (2000), Burnard and McEnery (2000), Aston (2001), Hunston (2002), Granger et al (2002), Tan (2002), Aston et al (2004) and Sinclair (2004). These works cover a wide range of issues related to using corpora in language pedagogy, e.g. corpus-based language description, corpus analysis in classroom, and learner corpora (cf. Keck 2004).

In the remainder of this article, we will explore the use of corpora in a number of areas in language pedagogy, including syllabus design, materials development, using corpora in classroom, domain-specific language and professional communication, language testing, teacher training, as well as learner corpus research.

## **2. Corpora and language pedagogy: a convergence**

In the opening chapter of *Teaching and Language Corpora* (Wichmann et al 1997), Leech observed that a convergence between teaching and language corpora was apparent. That convergence has three focuses, as noted by Leech (1997): the direct use of corpora in teaching (teaching about, teaching to exploit, and exploiting to teach), the indirect use of corpora in teaching (reference publishing, materials development, and language testing), and further teaching-oriented corpus development (LSP corpora, L1 developmental corpora and L2 learner corpora). These three focuses of convergence are worthy of note.

Of these focuses, 'teaching about' means teaching corpus linguistics as an academic subject like other sub-disciplines of linguistics such as sociolinguistics or discourse analysis. Corpus linguistics has now found its way into the curricula for linguistic and language related degree programs at both postgraduate and undergraduate levels. 'Teaching to exploit' means providing students with 'hands-on' know-how, as emphasized in McEnery et al (2005), so that they can exploit corpora for their own purposes. Once the student has acquired the necessary knowledge and techniques of corpus-based language study, learning activity may become student centred. If 'teaching about' is viewed as being associated typically with students of linguistics and languages, 'teaching to exploit' relates to students of all subjects which involve language study and learning. 'Exploiting to teach' means using a corpus-based approach to teaching language and linguistics courses, which would otherwise be taught using non-corpus-based methods. As for the indirect use of corpora in language teaching, we have already noted at the beginning of this article that corpora have revolutionized reference publishing in a manner such that people who have never heard of a corpus are using the products of corpus research. As we will see later

in the following sections, corpora also have a lot to offer in terms of syllabus design, materials development, classroom concordancing and language testing, etc. Finally, teaching-oriented corpora are particularly useful in teaching languages for specific purposes (LSP corpora) and in research on L1 (developmental corpora) and L2 (learner corpora) language acquisition, which will be discussed in section 8.

### **3. Syllabus design and materials development**

While corpora have been used extensively to provide more accurate descriptions of language use, a number of scholars have also used corpus data directly to look critically at existing TEFL (Teaching English as a Foreign Language) syllabuses and teaching materials. Mindt (1996), for example, finds that the use of grammatical structures in textbooks for teaching English differs considerably from the use of these structures in L1 English. He observes that one common failure of English textbooks is that they teach 'a kind of school English which does not seem to exist outside the foreign language classroom' (Mindt 1996: 232). As such, learners often find it difficult to communicate successfully with native speakers. A simple yet important role of corpora in language education is to provide more realistic examples of language usage. In addition, however, corpora may provide data, especially frequency data, which may further alter what is taught. For example, on the basis of a comparison of the frequencies of modal verbs, future time expressions and conditional clauses in corpora and their grading in textbooks used widely in Germany, Mindt (*ibid*) concludes that one problem with non-corpus-based syllabuses is that the order in which those items are taught in syllabuses 'very often does not correspond to what one might reasonably expect from corpus data of spoken and written English', arguing that teaching syllabuses should be based on empirical evidence rather than tradition and intuition with frequency of usage as a guide to priority for teaching (Mindt 1996: 245-246).

Hunston (2002: 189) echoes Mindt suggesting that 'the experience of using corpora should lead to rather different views of syllabus design.' The type of syllabus she discusses extensively is a 'lexical syllabus', originally proposed by Sinclair and Renouf (1988) and outlined fully by Willis (1990). According to Sinclair and Renouf (1988: 148), a lexical syllabus would focus on '(a) the commonest word forms in a language; (b) the central patterns of usage; (c) the combinations which they usually form.' While the term may occasionally be misinterpreted to indicate a syllabus consisting solely of vocabulary items, a lexical syllabus actually covers 'all aspects of language, differing from a conventional syllabus only in that the central concept of organization is lexis' (Hunston 2002: 189). Sinclair (2000: 191) would say that the grammar covered in a lexical syllabus is 'lexical grammar', not 'lexico-grammar', which attempts to 'build a grammar and lexis on an equal basis.' Indeed, as Murison-Bowie (1996: 185) observes, 'in using corpora in a teaching context, it is frequently difficult to distinguish what is a lexical investigation and what is a syntactic one. One leads to the other, and this can be used to advantage in a teaching/learning context.' Sinclair and his colleagues' proposal for a lexical syllabus is echoed by Lewis (1993, 1997a, 1997b, 2000) who provides strong support for the lexical approach to language teaching.

A focus of the lexical approach to language pedagogy is teaching collocations and the related concept of prefabricated units. There is a consensus that collocational knowledge is important for developing L1/L2 language skills (e.g. Bahns 1993; Zhang 1993; Cowie 1994; Herbst 1996: 389-391; Kita and Ogata 1997: 230-231; Partington 1998: 23-25; Hoey 2000, 2003; Shei and Pain 2000: 167-170; Sripicharn 2000: 169-170; Altenberg and Granger 2001; McEnery and Wilson 2001; McAlpine and Myles 2003: 71-75; Nesselhauf 2003). Hoey (2003), for example, posits that 'learning a lexical item entails learning what it occurs with and what grammar it tends to have.' Cowie (1994: 3168) observes that 'native-like proficiency of a language depends crucially on knowledge of a stock of prefabricated units.' Aston (1995) also notes that the use of prefabs can speed language processing in both comprehension and production, thus creating native-like fluency. A powerful reason for the employment of collocations, as Partington (1998: 20) suggests, 'lies in the way it facilitates communication processing on the part of hearer', because 'language consisting of a relatively high number of fixed phrases is generally more predictable than that which is not' while 'in real time language decoding, hearers need all the help they can get.' As such, competence in a language undoubtedly seems to involve collocational knowledge (cf. Herbst 1996: 389). Collocational knowledge indicates which lexical items co-occur frequently with others and how they combine within a sentence. Such knowledge is evidently more important than individual words themselves (cf. Kita and Ogata 1997: 230) and is needed for effective sentence generation (cf. Smadja and McKeown 1990). Zhang (1993), for example, finds that more proficient L2 writers use significantly more collocations, more accurately and in more variety than less proficient learners. Collocational error is a common type of error for learners (cf. McAlpine and Myles 2003: 75). Gui and Yang (2002: 48) observe, on the basis of the Chinese Learner English corpus, that collocation error is one of the major error types for Chinese learners of English. Altenberg and Granger (2001) and Nesselhauf (2003) find that even advanced learners of English have considerable difficulties with collocation. One possible explanation is that learners are deficient in 'automation of collocations' (Kjellmer 1991). 'As a result, learners need detailed information about common collocational patterns and idioms; fixed and semi-fixed lexical expressions and different degrees of variability; relative frequency and currency of particular patterns; and formality level' (McAlpine and Myles 2003: 75). Corpora are useful in this respect, not only because collocations can only reliably be measured quantitatively, but also because the KWIC (key word in centre) view of corpus data exposes learners to a great deal of authentic data in a structured way. Our view is line with Kennedy (2003), who discusses the relationship between corpus data and the nature of language learning, focusing on the teaching of collocations. The author argues that second or foreign language learning is a process of learning 'explicit knowledge' with awareness, which requires a great deal of exposure to language data.

#### **4. Classroom concordancing: data-driven learning**

While syllabus design and materials development are closely associated with what to teach, corpora have also provided valuable insights into how to teach. The issue of how to use corpora in the language classroom has been discussed extensively in the

literature. With the corpus-based approach to language pedagogy, the traditional ‘three P’s’ (Presentation – Practice – Production) approach to teaching may not be entirely suitable. Instead, the more exploratory approach of ‘three I’s’ (Illustration – Interaction – Induction) may be more appropriate, where ‘illustration’ means looking at real data, ‘interaction’ means discussing and sharing opinions and observations, and ‘induction’ means making one’s own rule for a particular feature (see Carter and McCarthy 1995: 155).<sup>4</sup> While the ‘three I’s’ approach was originally proposed by Carter and McCarthy (*ibid*) to teach spoken grammar, it may also apply to language education as a whole, in our view.

It is certainly clear that the teaching approach focusing on ‘three I’s’ is in line with Johns’ (1991) concept of ‘data-driven learning (DDL)’. Johns was perhaps among the first to realize the potential of corpora for language learners (e.g. Higgins and Johns 1984). In his opinion, ‘research is too serious to be left to the researchers’ (Johns 1991: 2). As such, he argues that the language learner should be encouraged to become ‘a research worker whose learning needs to be driven by access to linguistic data’ (*ibid*). Data-driven learning can be either teacher-directed or learner-led (i.e. discovery learning) to suit the needs of learners at different levels, but it is basically learner-centred. This autonomous learning process ‘gives the student the realistic expectation of breaking new ground as a “researcher”, doing something which is a unique and individual contribution’ (Leech 1997: 10).

Johns (1991) identifies three stages of inductive reasoning with corpora in the DDL approach: observation (of concordanced evidence), classification (of salient features) and generalization (of rules). The three stages roughly correspond to Carter and McCarthy’s (1995) ‘three I’s’. The DDL approach is fundamentally different from the ‘three P’s’ approach in that the former is bottom-up induction whereas the latter is top-down deduction. The direct use of corpora and concordancing in the language classroom has been discussed extensively in the literature (e.g. Tribble 1991, 1997a, 1997b, 2000, 2003; Tribble and Jones 1990, 1997; Flowerdew 1993; Karpati 1995; Kettemann 1995, 1996; Wichmann 1995; Woolls 1998; Aston 2001; Osborne 2001),<sup>5</sup> covering a wide range of issues including, for example, underlying theories, methods and techniques, and problems and solutions.

## **5. Languages for specific purposes and professional communication**

In addition to teaching English as a second or foreign language in general, a great deal of attention has been paid to domain-specific language use and professional communication (e.g. English for specific purposes and English for academic purpose). For example, Thurstun and Candlin (1997, 1998) explore the use of concordancing in teaching writing and vocabulary in academic English. Hyland (1999) compares the features of the specific genres of metadiscourse in introductory course books and research articles on the basis of a corpus consisting of extracts from 21 university textbooks for different disciplines and a similar corpus of research articles. Upton and Connor (2001) undertake a ‘moves analysis’ in the business English using a business learner corpus. The authors approach the cultural aspect of professional communication by comparing the ‘politeness strategies’ used by learners from different cultural backgrounds. Thompson and Tribble (2001) examine citation practices in academic text. Koester (2002) argues, on the basis of an analysis of the

performance of speech acts in workshop conversations, for a discourse approach to teaching communicative functions in spoken English. Yang and Allison (2003) study the organizational structure in research articles in applied linguistics. Carter and McCarthy (2004) explore, on the basis of the CANCODE corpus, a range of social contexts in which creative uses of language are manifested. Hinkel (2004) compares the use of tense, aspect and the passive in L1 and L2 academic texts. Xiao (2003) reviews a number of case studies using domain specialized multilingual corpora to teach domain specific translation. Studies such as these demonstrate that LSP corpora are particularly useful in teaching language for specific purposes and professional communication.

## **6. Language testing**

Another emerging area of language pedagogy which has started to use the corpus-based approach is language testing. Alderson (1996) envisaged the possible uses of corpora in this area: test construction, compilation and selection, test presentation, response capture, test scoring, and calculation and delivery of results. He concludes that '[t]he potential advantages of basing our tests on real language data, of making data-based judgments about candidates' abilities, knowledge and performance are clear enough. A crucial question is whether the possible advantages are born out in practice' (Alderson 1996: 258-259). The concern raised in Alderson's conclusion appears to have been addressed satisfactorily. Choi, Kim and Boo (2003), for example, find that computer-based tests are comparable to paper-based tests. A number of corpus-based studies of language testing have been reported. For example, Coniam (1997) demonstrated how to use word frequency data extracted from corpora to generate cloze tests automatically. Kaszubski and Wojnowska (2003) presented a corpus-driven program for building sentence-based ELT exercises – TestBuilder. The program can process raw and part-of-speech tagged corpora, tagged on the fly by a built-in part-of-speech tagger, and uses this as input for test material selection. Indeed, corpora have recently been used by major providers of test services for a number of purposes: 1) as an archive of examination scripts; 2) to develop test materials; 3) to optimize test procedures; 4) to improve the quality of test marking; 4) to validate tests; and 5) to standardize tests (cf. Ball 2001; Hunston 2002: 205). For example, the University of Cambridge Local Examinations Syndicate (UCLES) is active in both corpus development (e.g. Cambridge Learner Corpus, Cambridge Corpus of Spoken English, Business English Text Corpus and Corpus YLE Speaking Tests) and the analysis of native English corpora and learner corpora. At UCLES, native English corpora such as the British National Corpus (BNC) are used 'to investigate collocations, authentic stems and appropriate distractors which enable item writers to base their examination tasks on real texts' (Ball 2001: 7); the corpus-based approach is used to explore 'the distinguishing features in the writing performance of EFL/ESL learners or users taking the Cambridge English examinations' and how to incorporate these into 'a single scale of bands, that is, a common scale, describing different levels of L2 writing proficiency' (Hawkey 2001: 9); corpora are also used for the purpose of speaking assessment (Ball and Wilson 2002; Taylor 2003) and to develop domain-specific (e.g. business English) wordlists for use in test materials (Ball 2002; Horner and Strutt 2004).

## **7. Teacher training**

For learners to benefit from the use of corpora, language teachers must first of all be equipped with a sound knowledge of the corpus-based approach. It is unsurprising to discover then that corpora have been used in training language teachers (e.g. Allan 1999, 2002; Conrad 1999; Seidlhofer 2000, 2002; O’Keeffe and Farr 2003). Allan (1999), for example, demonstrates how to use corpus data to raise the language awareness of English teachers in Hong Kong secondary schools. Conrad (1999) presents a corpus-based study of linking adverbials (e.g. *therefore* and *in other words*), on the basis of which she suggests that it is important that a language teacher do more than using classroom concordancing and lexical or lexico-grammatical analyses if language teaching is to take full advantage of the corpus-based approach. Conrad’s concern with teacher education is echoed by O’Keeffe and Farr (2003), who argue that corpus linguistics should be included in initial language teacher education so as to enhance teachers’ research skills and language awareness.

## **8. Learner corpora and interlanguage analysis**

One of the most exciting recent developments in corpus-based language studies has been the creation and use of learner corpora in language pedagogy and interlanguage studies. At the pre-conference workshop on learner corpora affiliated to the International Symposium of Corpus Linguistics 2003 held at the University of Lancaster, the workshop organizers, Tono and Meunier, observed that learner corpora are no longer in their infancy but are going through their nominal teenage years – they are full of promise but not yet fully developed. In language pedagogy, the implications of learner corpora have been explored for curriculum design, materials development and teaching methodology (cf. Keck 2004: 99). The interface between L1 and L2 materials has been explored. Meunier (2002), for example, argues that frequency information obtained from native speaker corpora alone is not sufficient to inform curriculum and materials design. Rather, ‘it is important to strike a balance between frequency, difficulty and pedagogical relevance. That is exactly where learner corpus research comes into play to help weigh the importance of each of these’ (Meunier 2002: 123). Meunier also advocates the use of learner data in the classroom, suggesting that exercises such as comparing learner and native speaker data and analyzing errors in learner language will help students to notice gaps between their interlanguage and the language they are learning. Interlanguage studies based on learner corpora which have been undertaken so far focus on what Granger (2002) calls ‘Contrastive Interlanguage Analysis (CIA)’, which compares learner data and native speaker data, or language produced learners from different L1 backgrounds. The first type of comparison typically aims to identify under or overuse of particular linguistic features in learner language while the second type aims to uncover L1 interference or transfer. In addition to CIA, learner corpora have also been used to investigate the order of acquisition of particular morphemes. Readers can refer to Granger et al (2002) for recent work in the use of learner corpora, and read Granger (2003) for a more general discussion of the applications of learner corpora such as the International Corpus of Learner English (ICLE).

## **9. Frequency and authenticity: pros and cons**

Before we close the discussion of using corpora in language education, it is appropriate to address some objections to the use of corpora in language learning and teaching. While frequency and authenticity are often considered two of the most important advantages of using corpora, they are also the locus of criticism from language pedagogy researchers. For example, Cook (1998: 61) argues that corpus data impoverishes language learning by giving undue prominence to what is simply frequent at the expense of rarer but more effective or salient expressions. Widdowson (1990, 2000) argues that corpus data is authentic only in a very limited sense in that it is de-contextualized (i.e. traces of texts rather than discourse) and must be re-contextualized in language teaching. It can also be argued that:

on the contrary, using corpus data not only increases the chances of learners being confronted with relatively infrequent instances of language use, but also of their being able to see in what way such uses are atypical, in what contexts they do appear, and how they fit in with the pattern of more prototypical uses. (Osborne 2001: 486)

This view is echoed by Goethals (2003: 424), who argues that ‘frequency ranking will be a parameter for sequencing and grading learning materials’ because ‘[f]requency is a measure of *probability* of usefulness’ and ‘[h]igh-frequency words constitute a core vocabulary that is useful above the incidental choice of text of one teacher or textbook author.’ Hunston (2002:194-195) observes that ‘[I]tems which are important though infrequent seem to be those that echo texts which have a high cultural value’, though in many cases ‘cultural salience is not clearly at odds with frequency.’ While frequency information is readily available from corpora, no corpus linguist has ever argued that the most frequent is most important. On the contrary, Kennedy (1998: 290) argues that frequency ‘should be only one of the criteria used to influence instruction’ and that ‘[t]he facts about language and language use which emerge from corpus analyses should never be allowed to become a burden for pedagogy’. As such, raw frequency data is often adjusted for use in a syllabus, as reported in Renouf (1987: 168). It would be inappropriate, therefore, for language teachers, syllabus designers, and materials writers to ignore ‘compelling frequency evidence already available’, as pointed out by Leech (1997: 16), who argues that:

Whatever the imperfections of the simple equation ‘most frequent’ = ‘most important to learn’, it is difficult to deny that frequency information becoming available from corpora has an important empirical input to language learning materials.

If we leave objections to frequency data to one side, Widdowson (1990, 2000) also questions the use of authentic texts in language teaching. In his opinion, authenticity of language in the classroom is ‘an illusion’ (1990: 44) because even though corpus data may be authentic in one sense, its authenticity of purpose is destroyed by its use with an unintended audience of language learners (see Murison-Bowie 1996: 189). The implication of Widdowson’s argument is that only language produced for imaginary situations in the classroom is ‘authentic’. However, as argued by Fox (1987), invented examples often do not reflect nuances of usage. That is perhaps why, as Mindt (1996: 232) observes, students who have been taught ‘school



English' cannot readily cope with English used by native speakers in real life. As such, Wichmann (1997: xvi) argues that in language teaching, 'the preference for "authentic" texts requires both learners and teachers to cope with language which the textbooks do not predict.'

## 10. Conclusion

This article explored the potential uses of corpora in language education through a series of studies undertaken so far, focusing on TEFL. In conclusion, it is our view that corpora will not only revolutionize the teaching of subjects such as grammar in the 21<sup>st</sup> century (see Conrad 2000), they will also fundamentally change the ways we approach language education, including both what is taught and how it is taught. As Gavioli and Aston (2001) argue, corpora should not only be viewed as resources which help teachers to decide what to teach, they should also be viewed as resources from which learners may learn directly.

### Notes:

Robert de Beaugrande's unpublished paper, 'Large corpora and applied linguistics: H. G. Widdowson versus J. McH. Sinclair' (available online at <http://beaugrande.bizland.com/WiddowSincS.htm>), provides an excellent summary of the debate between Sinclair and Widdowson, at the Georgetown University Round Table on Languages and Linguistics in 1991, over the use of corpora in language teaching. While Widdowson, Sinclair and de Beaugrande characterize two extreme attitudes towards corpora, there are many milder (positive or negative) reactions to corpus data between the two extremes. Readers can refer to Nelson (2000: section 5.3.3.) for a good review.

<sup>2</sup> This article is based on Unit 10.8 of McEnery, Xiao and Tono (2005).

<sup>3</sup> The six TaLC conferences were held in Granada, Spain in July 2004; Bertinoro, Italy in July 2002; Graz, Austria in July 2000; Oxford, UK in July 1998; Lancaster, UK in August 1996 and Lancaster, UK in April 1994.

<sup>4</sup> Carter and McCarthy (1995: 155) comment that such a rule 'will be refined and honed as more and more data is encountered.' This progressive induction approach is what Murison-Bowie (1996: 191) would call the interlanguage approach: partial and incomplete generalizations are drawn from limited data as a stage on the way towards a fully satisfactory rule.

<sup>5</sup> See Tim John's website (<http://web.bham.ac.uk/johnstf/biblio.htm>) for a more comprehensive bibliography of classroom concordancing and data-driven learning.

### References:

- Alderson, C. 1996. 'Do corpora have a role in language assessment?' in J. Thomas and M. Short (eds.) *Using Corpora for Language Research*, pp. 248-259. London: Longman.
- Allan, Q. 1999. 'Enhancing the language awareness of Hong Kong teachers through corpus data'. *Journal of Technology and Teacher Education* 7/1: 57-74.

- Allan, Q. 2002. 'The TELEC secondary learner corpus: a resource for teacher development' in S. Granger, J. Hung and S. Petch-Tyson (eds.) *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*, pp. 195–212. Philadelphia: John Benjamins.
- Altenberg, B. and Granger, S. 2001. 'The grammatical and lexical patterning of MAKE in native and non-native student writing.' *Applied Linguistics* 22/2: 173-95.
- Aston, G. 1995. 'Corpora in language pedagogy: matching theory and practice' in G. Cook and B. Seidlhofer (eds.) *Principle and Practice in Applied Linguistics: Studies in Honour of H. G. Widdowson*. Oxford: Oxford University Press.
- Aston, G. (ed.) 2001. *Learning with Corpora*. Houston, TX: Athelstan.
- Aston, G, Bernardini, S. and Stewart, D. (eds.) 2004. *Corpora and Language Learners*. Amsterdam: John Benjamins.
- Bahns, J. 1993. 'Lexical collocations: a contrastive view'. *ELT Journal* 47/1: 56-63.
- Ball, F. 2001. 'Using corpora in language testing'. *Research Notes* 6: 6-8.
- Ball, F. 2002. 'Developing wordlists for BEC'. *Research Notes* 8: 10-13.
- Ball, F. and Wilson, J. 2002. 'Research projects relating to YLE Speaking Tests'. *Research Notes* 7: 8-10.
- Biber, D., Conrad, S. and Reppen, R. 1998. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- Biber, D., Johansson S., Leech G., Conrad S. and Finegan, E. 1999. *Longman Grammar of Spoken and Written English*. London: Longman.
- Burnard, L. and McEnery, A. (eds.) 2000. *Rethinking Language Pedagogy from a Corpus Perspective*. New York: Peter Lang.
- Carter, R. and McCarthy, M. 1995. 'Grammar and the spoken language'. *Applied Linguistics* 16/2: 141-158.
- Carter, R. and McCarthy, M. 2004. 'Talking, creating: interactional language, creativity, and context'. *Applied Linguistics* 25/1: 62-88.
- Choi, I., Kim, K. and Boo, J. 2003. 'Comparability of a paper-based language test and a computer-based language test'. *Language Testing* 20/3: 295–320.
- Coniam, D. 1997. 'A preliminary inquiry into using corpus word frequency data in the automatic generation of English language cloze tests'. *CALICO Journal* 16/2-4: 15-33.
- Conrad, S. 1999. 'The importance of corpus-based research for language teachers'. *System* 27: 1-18.
- Conrad, S. 2000. 'Will corpus linguistics revolutionize grammar teaching in the 21<sup>st</sup> century?'. *TESOL Quarterly* 34: 548–60.
- Cook, G. 1998. 'The uses of reality: a reply to Ronald Cater.' *ELT Journal* 52/1: 57-64.
- Cowie, A. 1994. 'Phraseology' in R. Asher (ed.) *The Encyclopaedia of Language and Linguistics* Vol. 6, pp. 3168-3171. Oxford: Pergamon Press Ltd.
- de Beaugrande, R. 2001. 'Interpreting the discourse of H. G. Widdowson: a corpus-based critical discourse analysis'. *Applied Linguistics* 22/1: 104-121.
- Flowerdew, J. 1993. 'Concordancing as a tool in course design'. *System* 21/3: 231-243.

- Fox, G. 1987. 'The case for examples' in J. Sinclair (ed.) *Looking Up: An Account of the COBUILD Project*, pp. 137-149. London: HarperCollins.
- Francis, G., Hunston, S. and Manning, E. 1996. *Collins COBUILD Grammar Patterns 1: Verbs*. London: HarperCollins.
- Francis, G., Hunston, S. and Manning, E. 1998. *Collins COBUILD Grammar Patterns 2: Nouns and Adjectives*. London: HarperCollins.
- Gavioli, L. and Aston, G. 2001. 'Enriching reality: language corpora in language pedagogy'. *ELT Journal* 55/3: 238-246.
- Goethals, M. 2003. 'E.E.T.: the European English Teaching vocabulary-list' in B. Lewandowska-Tomaszczyk (ed.) *Practical Applications in Language and Computers*, pp. 417-427. Frankfurt: Peter Lang.
- Granger, S. 2002. 'A bird's-eye view of learner corpus research' in S. Granger, J. Hung and S. Petch-Tyson (eds.) *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*, pp. 3-33. Philadelphia: John Benjamins.
- Granger, S. 2003. 'Practical applications of learner corpora' in B. Lewandowska-Tomaszczyk (ed.) *Practical Applications in Language and Computers*, pp. 291-302. Frankfurt: Peter Lang.
- Granger, S., Hung, J. and Petch-Tyson, S. (eds.) 2002. *Computer Learner Corpora, Second Language Acquisition, and Foreign Language Teaching*. Philadelphia: John Benjamins.
- Gui, S. and Yang, H. 2002. *Zhongguo Xuexizhe Yingyu Yuliaoku (Chinese Learner English Corpus)*. Shanghai: Shanghai Foreign Language Education Press.
- Herbst, T. 1996. 'What are collocations: sandy beaches or false teeth?'. *English Studies* 04/1996: 379-393.
- Higgins, J. and Johns, T. 1984. *Computers in Language Learning*. Oxford: Oxford University Press.
- Hinkel, E. 2004. 'Tense, aspect the passive voice in L1 and L2 academic texts'. *Language Teaching Research* 8/1: 5-29.
- Hoey, M. 2000. 'A world beyond collocation: new perspectives on vocabulary teaching' in M. Lewis (ed.) *Teaching Collocations*, pp. 224-245. Hove: Language Teaching Publications.
- Hoey, M. 2003. 'Lexical priming and the properties of text'. URL: [www.monabaker.com/tsresources/LexicalPrimingandthePropertiesofText.htm](http://www.monabaker.com/tsresources/LexicalPrimingandthePropertiesofText.htm).
- Horner, D. and Strutt, P. 2004. 'Analyzing domain-specific lexical categories: evidence from the BEC written corpus'. *Research Notes* 15: 6-8.
- Hunston, S. 2002. *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- Hyland, K. 1999. 'Talking to students: metadiscourse in introductory coursebooks'. *English for Specific Purposes* 18/1: 3-26.
- Johns, T. 1991. "'Should you be persuaded?": two samples of data-driven learning materials' in T. Johns and P. King (eds.) *Classroom Concordancing ELR Journal* 4. University of Birmingham.
- Karpati, I. 1995. *Concordance in Language Learning and Teaching*. Pecs: University of Pecs.

- Kaszubski, P. and Wojnowska, A. 2003. 'Corpus-informed exercises for learners of English: the TestBuilder program' in E. Oleksy and B. Lewandowska-Tomaszczyk (eds.) *Research and Scholarship in Integration Processes: Poland - USA – EU*, pp. 337-354. Łódź: Łódź University Press.
- Keck, C. 2004. 'Corpus linguistics and language teaching research: bridging the gap'. *Language Teaching Research* 8/1: 83-109.
- Kennedy, G. 1998. *An Introduction to Corpus Linguistics*. London: Longman.
- Kennedy, G. 2003. 'Amplifier collocations in the British National Corpus: implications for English language teaching'. *TESOL Quarterly* 37/3: 467-487.
- Kettemann, B. 1995. 'On the use of concordancing in ELT'. *TELL&CALL* 4: 4-15.
- Kettemann, B. 1996. 'Concordancing in English Language Teaching' in S. Botley, J. Glass, A. McEnery and A. Wilson (eds.) *Proceedings of Teaching and Language Corpora*, pp. 4-16. Lancaster University.
- Kettemann, B. and Marko, G. 2002. *Teaching and Learning by Doing Corpus Analysis*. Amsterdam: Rodopi.
- Kita, K. and Ogata, H. 1997. 'Collocations in language learning: corpus-based automatic compilation of collocations and bilingual collocation concordancer'. *Computer Assisted Language Learning* 10/3: 229-238.
- Kjellmer, G. 1991. 'A mint of phrases' in K. Aijmer and B. Altenberg (eds.) *English Corpus Linguistics: Studies in Honour of Jan Svartvik*. London: Longman.
- Koester, A. 2002. 'The performance of speech acts in workplace conversations and the teaching of communicative functions'. *System* 30: 167-184.
- Leech, G. 1997. 'Teaching and language corpora: a convergence' in A. Wichmann, S. Fligelstone, A. McEnery and G. Knowles (eds.) *Teaching and Language Corpora*, pp. 1-23. London: Longman.
- Lewis, M. 1993. *The Lexical Approach: The State of ELT and the Way Forward*. Hove: Language Teaching Publications.
- Lewis, M. 1997a. *Implementing the Lexical Approach: Putting Theory into Practice*. Hove: Language Teaching Publications.
- Lewis, M. 1997b. 'Pedagogical implications of the lexical approach' in J. Coady and T. Huckin (eds.) *Second Language Vocabulary Acquisition: A Rationale for Pedagogy*, pp. 255-270. Cambridge: Cambridge University Press.
- Lewis, M. (ed.) 2000. *Teaching Collocation: Further Developments in the Lexical Approach*. Hove: Language Teaching Publications.
- McAlpine, J. and Myles, J. 2003. 'Capturing phraseology in an online dictionary for advanced users of English as a second language: a response to user needs'. *System* 31: 71-84.
- McEnery, A. and Wilson, A. 2001. *Corpus Linguistics* (1<sup>st</sup> ed. 1996). Edinburgh: Edinburgh University Press.
- McEnery, A., Xiao, Z. and Tono, Y. 2005 (in press). *Corpus-based Language Studies: An Advanced Resource Book*. London: Routledge.
- Meunier, F. 2002. 'The pedagogical value of native and learner corpora in EFL grammar teaching' in S. Granger, J. Hung and S. Petch-Tyson (eds.) *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*, pp. 119–142. Philadelphia: John Benjamins.

- Mindt, D. 1996. 'English corpus linguistics and the foreign language teaching syllabus' in J. Thomas and M. Short (eds.) *Using Corpora for Language Research*, pp. 232-247. London: Longman.
- Murison-Bowie, S. 1996. 'Linguistic corpora and language teaching'. *Annual Review of Applied Linguistics* 16: 182-199.
- Nelson, M. 2000. A Corpus-Based Study of Business English and Business English Teaching Materials. PhD thesis, the University of Manchester, Manchester. Available at <http://www.kielikanava.com/thesis.html>.
- Nesselhauf, N. 2003. 'The use of collocations by advanced learners of English and some implications for teaching.' *Applied Linguistics* 24/2: 223-42.
- O'Keeffe, A. and Farr, F. 2003. 'Using language corpora in initial teacher education: pedagogic issues and practical applications'. *TESOL Quarterly* 37/3: 389-418.
- Osborne, O. 2001. 'Integrating corpora into a language-learning syllabus' in B. Lewandowska-Tomaszczyk (ed.) *PALC 2001: Practical Applications in Language Corpora*, pp. 479-492. Frankfurt: Peter Lang.
- Partington, A. 1998. *Patterns and Meanings*. Amsterdam: John Benjamins.
- Renouf, A. 1987. 'Moving on' in J. Sinclair (ed.) *Looking Up: An Account of the COBUILD Project*. London: HarperCollins.
- Seidlhofer, B. 2000. 'Operationalizing intertextuality: using learner corpora for learning' in L. Burnard and A. McEnery (eds.) *Rethinking Language Pedagogy from a Corpus Perspective*, pp. 207-24. New York: Peter Lang.
- Seidlhofer, B. 2002. 'Pedagogy and local learner corpora: working with learning driven data' in S. Granger, J. Hung and S. Petch-Tyson (eds.) *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*, pp. 213-34. Philadelphia: John Benjamins.
- Shei, C. and Pain, H. 2000. 'An ESL writer's collocational aid'. *Computer Assisted Language Learning* 13/2: 167-182.
- Sinclair, J. 2000. 'Lexical grammar'. *Naujoji Metodologija* 24: 191-203.
- Sinclair, J. (ed.) 2004. *How to Use Corpora in Language Teaching*. Amsterdam: John Benjamins.
- Sinclair, J. and Renouf, A. 1988. 'A lexical syllabus for language learning' in R. Carter and M. McCarthy (eds.) *Vocabulary and Language Teaching*. London: Longman.
- Sinclair, J., Bullon, S., Krishnamurthy, R., Manning, E. and Todd, J. 1990. *Collins COBUILD English Grammar*. London: HarperCollins.
- Smadja, F. and McKeown, K. 1990. 'Automatically extracting and representing collocations for language generation' in *Proceedings of the 28<sup>th</sup> Annual Meeting of Association for Computational Linguistics*, pp. 252-259.
- Sripicharn, P. 2000. 'Data-driven learning materials as a way to teach lexis in context' in C. Heffer, H. Sauntson and G. Fox (eds.) *Words in Context: A tribute to John Sinclair on his Retirement*. Birmingham: University of Birmingham.
- Stubbs, M. 2001. 'Texts, corpora, and problems of interpretation: a response to Widdowson'. *Applied Linguistics* 22/2: 149-172.
- Tan, M. 2002. *Corpus Studies in Language Education*. Bangkok: IELE Press.
- Taylor, L. 2003. 'The Cambridge approach to speaking assessment'. *Research Notes* 13: 2-4.

- Thompson, P. and Tribble, C. 2001. 'Looking at citations: using corpora in English for academic purposes'. *Language Learning & Technology* 5/3: 91-105.
- Thurstun, J. and Candlin, C. 1997. *Exploring Academic English: A Workbook for Student Essay Writing*. Sydney: NCELTR.
- Thurstun, J. and Candlin, C. 1998. 'Concordancing and the teaching of the vocabulary of academic English'. *English for Specific Purposes* 17: 267-280.
- Tribble, C. 1991. 'Concordancing and an EAP writing program'. *CAELL Journal* 1/2: 10-15.
- Tribble, C. 1997a. 'Corpora, concordances and ELT' in T. Boswood (ed.) *New Ways of Using Computers in Language Teaching*. Alexandria VA: TESOL.
- Tribble C. 1997b. 'Improving corpora for ELT: quick and dirty ways of developing corpora for language teaching' in B. Lewandowska-Tomaszczyk, P. Melia (eds.) *Practical Applications in Language Corpora – Proceedings of PALC '97*, pp. 107-117. Łódź: Łódź University Press.
- Tribble, C. 2000. 'Practical uses for language corpora in ELT' in P. Brett, and G. Motteram (eds.) *A Special Interest in Computers: Learning and Teaching with Information and Communications Technologies*, pp. 31-41. Kent: IATEFL.
- Tribble, C. 2003. 'The text, the whole text...or why large published corpora aren't much use to language learners and teachers' in B. Lewandowska-Tomaszczyk (ed.) *Practical Applications in Language and Computers*, pp. 303-318. Frankfurt: Peter Lang.
- Tribble, C. and Jones, G. 1990. *Concordances in the Classroom: A Resource Book for Teachers*. London: Longman.
- Tribble, C. and Jones, G. 1997. *Concordances in the Classroom: Using Corpora in Language Education*. Houston TX: Athelstan.
- Upton, T. and Connor, U. 2001. 'Using computerized corpus analysis to investigate the textlinguistic discourse move of a genre'. *English for Specific Purposes* 20: 313-329.
- Wichmann, A. 1995. 'Using concordances for the teaching of modern languages in higher education'. *Language Learning Journal* 11: 61-63.
- Wichmann, A. 1997. 'General introduction' in A. Wichmann, S. Fligelstone, A. McEnery and G. Knowles (eds.) *Teaching and Language Corpora*, pp. xvi-xvii. London: Longman.
- Wichmann, A. Fligelstone, S. McEnery A. and Knowles, G. (eds.) 1997. *Teaching and Language Corpora*. London: Longman.
- Widdowson, H. 1990. *Aspects of Language Teaching*. Oxford: Oxford University Press.
- Widdowson, H. 1991. 'The description and prescription of language' in J. Alatis (ed.) *Georgetown University Round Table on Languages and Linguistics 1991*, pp. 11-24. Washington, D.C.: Georgetown University Press.
- Widdowson, H. 2000. 'The limitations of linguistics applied'. *Applied Linguistics* 21/1: 3-25.
- Willis, D. 1990. *The Lexical Syllabus: A New Approach to Language Teaching*. London: HarperCollins.

- Woolls, D. 1998. 'Multilingual parallel concordancing for pedagogical use' in *Teaching and Language Corpora*, pp. 222-227. Keble College, Oxford, 24-27 July 1998.
- Xiao, Z. 2003. 'Use of parallel and comparable corpora in language study'. *English Education in China* 2003(1).
- Yang, Y. and Allison, D. 'Research articles in applied linguistics: moving from results to conclusions'. *English for Specific Purposes* 22: 365-385.
- Zhang, X. 1993. *English Collocations and Their Effect on the Writing of Native and Non-native College Freshmen*. PhD thesis. Indiana University of Pennsylvania.