

# **The Potential of Corpus-Informed L2 Pedagogy**

Jonathon Reinhardt  
University of Arizona

## Abstract

Corpus linguistic methods have led to many revelations about the nature of language use and language learning which are otherwise untenable. To date, however, these findings have not had major impact on L2 (second and foreign language) teaching, except in a few areas like L2 for specialized purposes. This may be due to a number of reasons, including issues of theoretical commensurability and the nature of ‘teaching revolutions’. Still, corpus continues to influence SLA researchers, and pedagogical innovators continue to design, implement and evaluate corpus-informed L2 curricula. The future of corpus in L2 pedagogy depends on the extent to which this research and innovation can impact professional L2 instructor education and the design of commercial L2 teaching materials, as well as the continuing development of corpora and corpus-informed resources that are accessible, diverse, and adaptable.

## **1. Introduction**

In the 1960s, when linguists like Kucera & Francis (1967) first began collecting large amounts language data to analyze with the huge mainframes of the day, Noam Chomsky had only recently challenged structural and behaviorist paradigms (1957) with transformational grammar. Chomsky argued that introspection and grammaticality judgments were the most effective means to gain insights into the nature of linguistic competence, implying that analysis of performance data was unnecessary, even pointless (see Toolan 2007). Kucera & Francis, however, put their trust in empirical analysis and developed the Brown corpus, and began discovering facts about language use that introspection by ‘armchair linguists’ could never reveal.

As computer power increased through the next few decades, corpus linguistics became more widespread, and its discoveries more profound. Researchers found words behave in chunks and collocate with each other at varying strengths. Words colligate with grammatical constructions so much so that the distinction between the lexicon and grammar is blurred. Rules seem to be probabilistic rather than deterministic, and can be better described as tendencies or preferences rather than fixed rules. The lexicogrammatical profile of a text, including its most frequent

words, chunks, and constructions, is so influenced by its context of use that registers can be understood to have their own grammars (Aston 2001). The distinction is especially great between written and spoken language.

As corpus linguists made these observations, many applied linguists took note and began shifting their research practices. Most successfully, lexicographers and test designers have transformed their fields using corpus-based approaches, and today most English dictionaries and standardized tests (e.g. TOEFL) are based on corpus findings, as well as those in many other languages.<sup>1</sup> Scholars of first and second language acquisition have found in corpus-based techniques a powerful method for the analysis of language learning and development. The areas least changed are L2 (second and foreign language) teacher training, instructional methods, and materials design, which have yet to incorporate or reflect the implications of corpus-informed linguistic research by any significant measure. Except for a few areas, namely advanced ELT (English language teaching, as a second or foreign language) and L2 for specific purposes, corpus-informed L2 teaching has not spread far.

In the first issue of *Studies in Hispanic and Lusophone Linguistics*, Mark Davies, the creator of *Corpus del Español* ([www.corpusdelespanol.org](http://www.corpusdelespanol.org)) and the *Corpus do Português* ([www.corpusdoportugues.org](http://www.corpusdoportugues.org)) provides an excellent overview of corpus linguistics in Spanish and Portuguese (Davies 2008). Davies provides a summary of the conflicts between Chomskyan approaches and corpus-based linguistic theory, and explains that many early doubts dissolved once corpora became statistically large enough to alleviate criticism. He also outlines the basics of corpus analysis, presents many different corpora and related resources for the study of Spanish and Portuguese, and discusses what researchers might do with the corpora he developed. He concludes with a brief survey of corpus-based research in the two languages.

Because of its comprehensiveness, I refer the reader to Davies' article for a thorough discussion of corpus design and the possibilities of general research on Spanish and Portuguese using corpus resources and methodologies. As a complement to that paper, which only very briefly touches on pedagogy, I will focus on the pedagogical application of corpus-based theory and methodologies in this *Viewpoints* paper, and explore possible reasons why current L2 teaching practices generally do not reflect the knowledge we have gained through corpus linguistics. To begin, I'll outline early debates and discussions, most of which occurred in ELT, then present a brief overview of corpus findings about the nature of language and learning. A discussion of corpus-informed L2 pedagogy will follow, and I'll conclude with ideas on how corpus linguistics might make more of an impact on L2 teaching in the future.

## 2. Early debates and discussions

Since corpus-informed techniques were first brought into the L2 classroom, a variety of pedagogical issues have been touted as serious impediments to their widespread adoption. About a decade ago, there were several prominent discussions in journals and at conferences among applied linguists about the role of corpus linguistics in ELT. At the time, the largest corpora were of English, and there had been more corpus-based research on English than on other languages. While earlier studies had used corpus techniques to research English lexicogrammar, researchers were also beginning to apply corpus-informed approaches to the analysis of learner language and development. The conversations are worth recounting, as they provide insight into why corpus-informed L2 teaching has not become more widespread.

The discussions focused on questions of ‘real language’, authenticity, the native speaker, and the relationship between the linguist and the language teacher. Corpus linguists like John Sinclair and Michael Stubbs had taken the ‘strong’ position (McEnery, Xiao, & Tono 2006, p.131) that corpus analysis should provide the basis for all linguistic research, which would trickle down to language teaching. Seidlhofer (2003) describes this as a “kind of natural alliance between linguistic description and pedagogic prescription” (p.78) inferred by the strong position, which linguists like Henry Widdowson (1991) began to challenge by questioning the applicative value of corpus findings. Widdowson argued that as with any descriptive data, corpus data must be interpreted and made relevant to learner needs by considering aspects of learner development. Authentic as corpus data were, language learners needed to authenticate the data themselves in the process of becoming members of a linguistic community, a process necessarily mediated by teachers and materials designers.

In an early journal-based debate with Guy Cook, Carter (1998) argued that corpus findings on spoken English usage were radical enough to change curricula. For example, even though three-part exchanges were more common in actual spoken interactions, with the third turn being an often evaluative ‘chunk’, most ELT materials presented conversations as two-part exchanges. In addition, spoken language demonstrated far more vague language, ellipsis, and topic-fronted ‘heads’ (e.g. “now chocolate, I like”) among other features, than were represented in materials. Carter advocated language awareness pedagogy, informed by corpus-based research, where learners would be led to explicitly notice, discuss, and analyze linguistic features in relationship to the formal, functional, and interpersonal features of a given text. In this way, learners would be able to interact with the text meaningfully and authentically as critical observers (Gavioli & Aston 2001).

Cook’s (1998) reply to Carter urged measured reception to overzealous corpus and applied linguists “for whom corpus findings are the only source of truth” (p.57). Cook criticized the fact that all corpus words are treated equally, even though some words may be from a more widely received text (e.g. a political speech) than others,

and be received with more salience than others. He observed, “there is no straightforward correlation between the words people use, the intentions they had in them, and the interpretations which other people put upon them” (p.58), and maintained that corpora cannot and do not tell us everything there is to understand about language use. Cook also discussed the danger of holding the corpus up as a native speaker norm to emulate and “stressing what is actually done at the expense of what is appropriate in a particular context” (p.60). Finally, Cook echoed Widdowson in calling for pedagogical mediation of corpora, and chastened those who believe that a linguistic revolution “necessarily constitutes a pedagogic one” (p.62).

As Cook noted, another criticism of the use of corpora in L2 teaching and research revolves around the fact that most corpora are collections of native speaker uses, and therefore present the native speaker as the ideal (Prodromou 1996). Because native speaker status is never actually attainable for the learner, and the desirability of native-like use is only assumed and not confirmed, this perspective is seen to disempower learners and reduce their agency in the learning process. While more recent corpus-informed L2 pedagogy incorporate discovery and language awareness approaches, where language use is presented as choice rather than prescription, implications remain any time that native production is used as a norm. Seen critically, this is a broader issue for all L2 teaching, corpus-informed or not.

### **3. The nature of language and language learning**

Although the above discussions may have faded, they still resonate and may offer reasons why corpus still has not made headway in L2 teaching: 1) that corpus linguists are over exaggerating the revolutionary quality of their findings, 2) that corpora do not capture essential aspects of language use, 3) that the language captured in corpora is devoid of context and thus inauthentic, and 4) that corpus-based pedagogy can be just as prescriptive as other approaches. While scholars and curriculum designers have attempted to address these in their instructional innovations, it is the first claim that deserves attention, because of its paradigm-shifting implications.

There is a fundamental debate on whether corpus linguistics is a theoretically neutral methodology (cf. McEnery & Wilson 2001), or whether it entails a particular orientation towards language and research (Sinclair 1991 and Stubbs 1996). Sinclair (1991) and others use the term ‘phraseology’ to refer to the corpus-informed conceptualization of language as prefabricated chunks or formulaic sequences that exhibit collocational (word-to-word) and colligational (word-to-syntax) tendencies. Meaning can be associated with a grammatical structure or construction, not just the individual words themselves, so that particular chunks carry ‘semantic prosody’. Sinclair has termed this approach the ‘idiom principle’ (1991), where each word in a text “is used in a common phraseology, meaning is

attached to the whole phrase rather than the individual parts of it, and the hearer or reader understands the phrase as a phrase rather than as a grammatical template with lexical items in it" (Hunston 2003, p.143). The idiom principle is balanced by the open-choice principle, where lexicogrammatical choice is less fixed; in other words, words and constructions exhibit a sort of attraction to their neighbors to a greater (idiomatic) or lesser (open) degree. Furthermore, there is evidence from corpus-based research for the psycholinguistic reality of the idiom principle (Schmitt, Grandage & Adolphs 2004 and Erman 2007). This understanding of the nature of language is, in many minds, incommensurable with the Chomskyan view that language is modular, i.e. syntax and semantics are separate, and empirical data have no value for theorizing. Chomsky has said that "grammar is autonomous and independent of meaning" (1957), and more recently, famously or infamously depending on one's philosophy, that "corpus linguistics doesn't mean anything" (Andor 2004).

Whether or not Chomskyan concepts like universal grammar have implications for language teaching (see White 2003), corpus methodologies are having a great impact on cognitive, constructionist and usage-based theories of first and second language acquisition (Ellis 2003, Goldberg 1995, Langacker 1987 and Tomasello 2003). Constructionist theories generally reject the notion of a language acquisition device, and consider language learning to follow "general cognitive principles of category learning, with schematic constructions emerging from usage" (Ellis & Collins 2009, p.329), and so rely primarily on usage data. Using corpus methods, research has found that the learning of a construction may be strongly affected by its frequency and recency, as well as its salience, prototypicality, and functional redundancy. These findings have implications for what particular language is presented to L2 learners, in what ways, and in what order.

Research using L2 learner corpora, i.e. corpora comprised of L2 learner language, has also contributed to our understanding of language learning and development (cf. Nesselhauf 2004). Granger (1998) advocates contrastive and error analytic techniques to compare features of learner language and expert or native speaker production from genre-compatible corpora. Petch-Tyson (1998) and later Cobb (2003), for example, found that English learners used interpersonal involvement markers in writing, like pronouns, emphatic particles, and questions, much more frequently than native speakers. Belz & Vyatkina (2005) used a learner corpus to trace how their students developed the pragmatic competence to use German discourse particles appropriately in telecollaborative exchanges. These methods are becoming more widespread, as corpus resources and methods become more accessible to researchers.

#### 4. Corpus-informed L2 pedagogical innovations

Corpus-informed pedagogy has been gaining some proponents, particularly in advanced L2 for specific purposes. Tim Johns (1991) pioneered the use of corpus techniques in the L2 classroom, an approach he termed data-driven learning (DDL). In DDL, sorted concordances of words and phrases are presented to students, who induce meanings and identify form-function relationships and patterns of semantic prosody. Reflecting a strong phraseological orientation, the approach is corpus-driven, i.e. all lexicogrammatical understandings emerge from the corpus, and not from the analyst's preconceptions about grammar. DDL is expanded into an entire syllabus with Willis' Lexical Approach (1990), which considers the organizing concept to be the lexicon, rather than pre-determined grammatical points. Both approaches were criticized for being too radically corpus-driven. Critics pointed out that because each line of the concordance is from a different corpus or part of the corpus, contextual data is less discernable, and because each line is given equal weight, the salience of that particular usage is unknown. In practice, students felt overwhelmed by the huge number of seemingly random language bits and demands for inductive thinking.

The need to contextualize corpus data and incorporate language awareness principles became more apparent, and DDL was conceptualized more broadly. Reflecting this understanding, Gabel (2001) states that DDL can "foster the learners' analytical capacities, promote their explicit knowledge of the L2, facilitate critical language awareness, and support the development of learner autonomy" (p.269). Gavioli (1997) argues that corpus-informed approaches like DDL necessarily involve training the students to develop a researcher mindset, or else risks overwhelming them. Students should be introduced gradually to analytic methods like text reconstruction, and encouraged to make hypotheses based on observed patterns. Once a hypothesis is developed, the methods are used once again to confirm or refute it, leading to new hypotheses. Through practice exploring corpora over time, the student develops an autonomous, serendipitous attitude towards language learning (Bernardini 2003).

There has also been focus on promoting learners' awareness of their own usage. Seidlhofer (2002) advocates a 'learner-driven data' approach where students develop and analyze mini-corpora of their own language production, thus promoting the noticing of output (Swain 1995). Coniam (2004) explains how corpus-based methods can be used to develop individual textual profiles. Lee & Swales (2006) trained doctoral students in corpus analysis and had them compile corpora of their own writing, which the students were able to compare with academic writing from their fields. The authors claim their approach resulted in raised rhetorical consciousness by decentering the students' relationship with professional discourse.

Calls for pedagogical mediation are also being addressed. Braun (2005) argues that the features which make corpora and corpus-based methods useful for

researchers, namely size, content, design, and ‘messiness’ of results, make them difficult to use for language learners. She presents a solution for students with an online corpus tool structured around thematic videos and supported with corpus-based activities. Flowerdew (2005, 2009) maintains that pedagogical innovators are increasingly balancing bottom-up techniques like DDL with top-down approaches that incorporate discussion of context and user variables. Innovators are using smaller, more localized corpora that reflect the register-specific reality of particular textual features, and more corpora are being annotated for contextual information, including speaker, audience, and text purpose. Learner-friendly tools designed according to pedagogical principles are allowing scaffolded access to corpus-informed resources and better integration into L2 syllabi.

Finally, corpus-informed language teaching is being positively assessed and evaluated. Liu & Jiang (2009) combined a discourse-focused approach, a lexicogrammatical perspective, and DDL techniques to teach English grammar to EFL students. They found students responded positively with increased language awareness, including critical understanding of grammar and the importance of context. They recommend modeling DDL activities first, presenting deductive corpus searching before inductive techniques (i.e., testing a hypothesis based on a known lexicogrammatical rule), doing the activities in groups, and having learners use corpus techniques to verify the answers in their workbook’s exercise key.

## **5. Future potentials**

Innovations notwithstanding, corpus linguistics has not had much impact on the L2 teaching mainstream, with the exception of dictionaries and some textbooks like the Touchstone series (McCarthy, McCarten & Sandiford 2005) for ELT. For example, Römer (2005) surveyed popular ELT textbooks and found that English progressive forms are presented with more frequency than occur in natural corpora. Shortall (2007) noted that English present perfect is presented more frequently in textbooks with ‘yet’ and ‘already’ than occurs naturally, according to corpus analysis. There have been few similar studies of other languages, but an informal survey of commercially available L2 classroom texts shows traditional presentations of grammar and vocabulary, not informed by the findings of corpus linguistics by any significant measure.

In 2000, Susan Conrad noted that many important corpus-informed findings on the nature of lexicogrammar had not made much headway in ELT, but she still retained optimism for the new millennium. If the implications of corpus linguistics research were more widely accepted, she predicted, teachers and students would start to see grammar as register-specific rather than monolithic, grammar and vocabulary instruction would become more integrated, and contextual appropriateness would become more important than structural accuracy. She offered several suggestions towards effecting these changes that are applicable to all L2

pedagogy: 1) that corpus linguists should direct research to language teachers and interpret findings for L2 pedagogical implications, 2) that those findings should be presented as only part of the picture, not as panaceas, and 3) that they should inform and be integrated into, rather than drive or revolutionize, materials design. Finally, she notes that for real change, teachers themselves need to be willing to deviate from traditional approaches to grammar teaching.

A decade later, Conrad's suggestions still hold. The suggestion that corpus research should be more accessible can be addressed in several ways. First and foremost, corpus-informed knowledge and methods need to be integrated into introductory linguistics, SLA, teaching methods, and pedagogical grammar courses. It may be that corpus-informed approaches have not made their way into L2 classrooms because instructors have not been exposed to corpus-based findings or methods in their own education and professional development. Much research has shown that teaching involves application of a complex combination of content knowledge, pedagogical knowledge derived from education, and personal knowledge derived from experience (e.g. Golombek 1998). We can say we believe one thing about the nature of language, learning, and teaching, but our actual teaching may reflect something else—usually our experiences as teachers and learners. Still, formal training can have an effect, depending on the degree to which it is itself experiential and reflective.

While the reasons we teach as we do are complex, it is no secret that many L2 instructors, especially novices, teach 'from the book'. Corpus linguistics might be more influential in L2 instruction if it can influence the design of the materials with which teachers teach, especially if combined with training and professional development. To this end, corpus-based findings and understandings need to make their way to materials designers, leading them to confront their own assumptions about what language is, how learning happens, and how these are reflected in the materials they create. Usage-based research is challenging the natural order hypothesis and entrenched beliefs about which structures should be taught in what order, and how this should be done. Chunking tendencies, frequencies, salience, prototypicality, and register specificity should be considered in the teaching of vocabulary and grammar, that is, what should be conceptualized as lexicogrammar. These findings and understandings need to be pedagogically mediated and presented in forms that serve L2 pedagogical goals and objectives.

Corpus linguistics will also have more influence on L2 teaching if corpus-informed materials, corpus analysis tools, and well-designed corpora are simply more numerous, accessible and user friendly, and preferably web-based. As it stands now, many corpora and tools are expensive and difficult to use, except for a small number of web-based resources that integrate a corpus and tool and offer it for free, like Davies' *Corpus del Español* and the *Michigan Corpus of Academic Spoken English* ([quod.lib.umich.edu/m/micase/](http://quod.lib.umich.edu/m/micase/)). Other tools like Sabine Braun's *ELISA* ([www.uni-tuebingen.de/elisa/html/elisa\\_index.html](http://www.uni-tuebingen.de/elisa/html/elisa_index.html)), Tom Cobb's *Lextutor*

([www.lex tutor.ca](http://www.lex tutor.ca)), and Sharp Laboratories of Europe's *JustTheWord* (193.133.140.102/JustTheWord/), while for ELT, are good models for learner friendly tools based on corpus-informed pedagogical principles of any L2.

This leads to Conrad's second suggestion, that corpus-informed findings and approaches be integrated into pedagogical curricula and materials gradually, rather than be presented in isolation or as revolutionary. Mediation of materials, techniques, and curricula is a necessity. A class accustomed to traditional presentation of vocabulary may initially resist the inductive thinking demands of a concordance, but an instructor can prepare a concordance, selecting the lines to include beforehand, to be sure the instances presented are relevant and the contexts discernable. Contextualization of corpus data is crucial, and can be done, as suggested by Flowerdew (2009), by balancing bottom-up inductive methods with top-down, genre-focused instruction. If L2 instruction becomes more specialized, for example Spanish for medical purposes or Portuguese for business, corpus-informed pedagogy will be seen as a natural fit. Moreover, although corpus may seem to fit better with advanced proficiency curricula, corpus principles can still inform lower level curricula and materials—dialogues can reflect corpus-based findings, vocabulary can be presented with collocational and frequency information, and grammatical explanations can be contextualized in discussions of register appropriateness. Inductive analysis, active learning, and language awareness are not beyond lower proficiency learners' capabilities.

Conrad's third suggestion was that corpus-informed techniques should be combined with other approaches. While corpus-informed pedagogy fits with language awareness, discovery learning, and discourse-functional approaches, these approaches encompass more than just corpus, and may demand that instructors and students think of language and learning in new ways. With any application of corpus methods or findings, presenting language use as choice is important, as it can empower learners and balance the prescriptive message that relying on corpus-based data may imply. Moreover, these new approaches are commensurable with, and can be scaffolded into, more familiar approaches that focus on learning through meaningful language use and the development of critical thinking and autonomous learning skills. This also relates to Conrad's caveat, that to adopt corpus-informed techniques may necessitate yielding authority in the classroom to a corpus and to learner discoveries. In practice, corpus should be a natural fit with the orientation, already promoted by communicative and constructivist approaches, where instructors see themselves as guides to and facilitators of language learning, rather than as the ultimate sources of knowledge.

## 6. Conclusion

Corpus linguistics has produced many revelations about the nature of language use and language learning, but has failed to have the initially predicted revolutionary

impact on linguistics and L2 teaching. Early debates and the experiences of early adopters raise important pedagogical caveats and considerations regarding the use of corpus techniques in the L2 classroom that still hold true today. There may be issues of ontological and epistemological commensurability between a phraseological orientation towards language and a generative one, but they may be less important in L2 teaching practice, where the reasons why particular approaches are adopted, adapted, or ignored are complex.

However, as the findings and methods of corpus linguistics continue to impact SLA researchers, L2 teaching materials developers, and individual pedagogical innovators, a critical mass that tips the scale may be reached. As with the development of corpus linguistics itself, the catalyst may be technology—as L2 learning becomes more technology-mediated, individual learners and instructors may find that corpora and corpus-informed learning resources are increasingly more accessible, diverse, and adaptable. We can only hope that they continue to innovate for their own teaching, learning, and research needs, and that they share what they create, as have pioneers like Davies, Cobb, and Braun. Like corpus-informed L2 teaching, a corpus-informed L2 teaching revolution will have more of an impact if it is both bottom-up and top-down.

## Notes

1. Davies (2008) notes that the earliest Spanish language dictionary to include frequency data was Juilland & Chang-Rodríguez' *Frequency Dictionary of Spanish Words*, published in 1964.

## References

- Andor, József. 2004. The master and his performance: An interview with Noam Chomsky. *Intercultural Pragmatics* 1(1), 93–111.
- Aston, Guy. 2001. *Learning with corpora*. Houston: Athelstan.
- Belz, Julie & Nina Vyatkina. 2005. Learner corpus research and the development of L2 pragmatic competence in networked intercultural language study: The case of German modal particles. *Canadian Modern Language Review/Revue canadienne des langues vivantes* 62(1), 17-48.
- Bernardini, Sylvia. 2003. Exploring New Directions for Discovery Learning. In Bernhard Kettemann & Georg Markö (eds.), *Language and computers, teaching and learning by doing corpus analysis*, 165-182. Amsterdam: Rodopi
- Braun, Sabine. 2005. From pedagogically relevant corpora to authentic language learning contents. *ReCALL* 17(1), 47-64.
- Erman, Britt. 2007. Cognitive processes as evidence of the idiom principle. *International Journal of Corpus Linguistics* 12(1), 25–53.

- Carter, Ronald. 1998. Orders of reality: CANCODE, communication, and culture. *ELT Journal* 52(1), 43-56.
- Chomsky, Noam. 1957. *Syntactic structures*. The Hague: Mouton.
- Cobb, Tom. 2003. Analyzing late interlanguage with learner corpora: Québec replications of three European studies. *The Canadian Modern Language Review/La Revue canadienne des langues vivantes* 59, 393-423.
- Coniam, David. 2004. Concordancing oneself: Constructing individual textual profiles. *International Journal of Corpus Linguistics* 9, 271-298.
- Conrad, Susan. 2000. Will corpus revolutionize grammar teaching in the 21st century? *TESOL Quarterly* 34, 548-560.
- Cook, Guy. 1998. The uses of reality: A reply to Ronald Carter. *ELT Journal* 52(1) 57-63.
- Davies, Mark. 2008. New directions in Spanish and Portuguese corpus linguistics. *Studies in Hispanic and Lusophone Linguistics* 1, 149-186.
- Ellis, Nick & Laura Collins. 2009. Input and second language acquisition: The roles of frequency, form, and function. *The Modern Language Journal* 93, 329-335.
- Ellis, Nick. 2003. Constructions, chunking, and connectionism: The emergence of second language structure. In Catherine Doughty & Michael Long (eds.), *Handbook of second language acquisition*. Oxford: Blackwell.
- Flowerdew, Lynne. 2005. An integration of corpus-based and genre-based approaches to text analysis in EAP/ESP: Countering criticisms against corpus-based methodologies. *English for Specific Purposes* 24, 321-332.
- Flowerdew, Lynne. 2009. Applying corpus linguistics to pedagogy: A critical evaluation. *International Journal of Corpus Linguistics* 14, 393-417.
- Gabel, Stefan. 2001. Over-indulgence and under-representation in interlanguage: Reflections on the utilization of concordancers in self-directed foreign language learning. *Computer Assisted Language Learning* 14, 269-288.
- Gavioli, Laura & Guy Aston. 2001. Enriching reality: Language corpora in language pedagogy. *ELT Journal* 55, 238-246.
- Gavioli, Laura. 1997. Exploring texts through the concordancer: Guiding the learner. In Anne Wichmann, Steven Fligelstone, Tony McEnery, & Gerry Knowles (eds.), *Teaching and Language Corpora*, 83-99. London: Longman.
- Goldberg, Adele. 1995. *A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.
- Golombek, Paula. 1998. A case study of second language teachers' personal practical knowledge. *TESOL Quarterly* 32, 447-464.
- Granger, Sylviane. 2002. A Bird's-eye view of learner corpus research. In Sylviane Granger, Joseph Hung & Stephanie Petch-Tyson (eds.), *Computer learner corpora, second language acquisition and foreign language teaching*, 2-33. Amsterdam: John Benjamins.
- Hunston, Susan. 2002. *Corpora in applied linguistics*. Cambridge: Cambridge University Press.

- Johns, Tim. 1991. From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. *English Language Research Journal* 4, 27-45.
- Kucera, Henry & W. Nelson Francis. 1967. *Computational analysis of present-day American english*. Providence: Brown University Press.
- Langacker, Ronald. 1987. *Foundations of cognitive grammar: Theoretical prerequisites*. Stanford: Stanford University Press.
- Lee, David & John Swales. 2006. A corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora. *English for Specific Purposes* 25, 56-75.
- Liu, Dilin & Ping Jiang. 2009. Using a corpus-based lexicogrammatical approach to grammar instruction in EFL and ESL contexts. *The Modern Language Journal* 93(1), 61-78.
- McCarthy, Michael, Jeanne McCarten & Helen Sandiford. 2005. *Touchstone series*. Cambridge: Cambridge University Press.
- McEnery, Tony & Andrew Wilson. 2001. *Corpus linguistics*. Edinburgh: Edinburgh University Press.
- McEnery, Tony, Richard Xiao & Yukio Tono. 2006. *Corpus-based language studies*. London: Routledge.
- Nesselhauf, Nadia. 2004. Learner corpora and their potential for language teaching. In John Sinclair (ed.), *How to use corpora in language teaching* (Vol. 12), 125-156. Amsterdam: John Benjamins.
- Petch-Tyson, Stephanie. 1998. Writer/reader visibility in EFL written discourse. In Sylviane Granger (ed.), *Learner English on computer*, 107-118. New York: Longman.
- Prodromou, Luke. 1996. 'Correspondence'. *ELT Journal*, 50, 371-373.
- Römer, Ute. 2005. *Progressives, patterns, pedagogy: A corpus-driven approach to English progressive forms, functions, contexts and didactics*. Amsterdam: John Benjamins.
- Schmitt, Norbert, Sarah Grandage & Svenja Adolphs. 2004. Are corpus-derived recurrent clusters psycholinguistically valid? In Norbert Schmitt (ed.), *Formulaic sequences*, 127-149. Amsterdam: John Benjamins.
- Seidlhofer, Barbara. 2002. Pedagogy and local learner corpora: Working with learning-driven data. In Sylviane Granger, Joseph Hung & Stephanie Petch-Tyson (eds.), *Computer learner corpora, second language acquisition and foreign language teaching*, 213-233. Amsterdam: John Benjamins.
- Seidlhofer, Barbara. 2003. *Controversies in applied linguistics*. Oxford: Oxford University Press.
- Shortall, Terry. 2007. The L2 syllabus: Corpus or contrivance? *Corpora* 2(2), 157-185.
- Sinclair, John. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

- Stubbs, Michael. 1996. *Text and corpus analysis: Computer assisted studies of language and culture*. Oxford: Blackwell.
- Swain, Merrill. 1995. Three functions of output in second language learning. In Guy Cook & Barbara Seidlhofer (eds.), *Principles and practice in the study of language*. Oxford: Oxford University Press.
- Tomasello, Michael. 2003. *Constructing a language: A usage-based theory of language acquisition*. Cambridge: Harvard University Press.
- Toolan, Michael. 2007. Trust and text, text as trust. *International Journal of Corpus Linguistics* 12, 269-288.
- White, Lydia. 2003. *Second language acquisition and universal grammar*. Cambridge: Cambridge University Press.
- Widdowson, Henry. 1991. The description and prescription of language. In James Alatis (ed.): *Linguistics and language pedagogy: The state of the art*. Washington, DC: Georgetown University Press.
- Willis, Dave. 1990. *The lexical syllabus: A new approach to language teaching*. London: Collins ELT.