Integrating Corpora in Language Learning and Teaching

ANGELA CHAMBERS
University of Limerick
(*email: Angela.Chambers@ul.ie*)

The contribution of corpora to the language-learning environment has been developing and gaining momentum in the course of the last 50 years. According to McEnery and Wilson (1997:12), the first attested use of concordances in the classroom was as early as 1969, by Peter Roe at Aston University. An important milestone in the 1970s was the publication of Sinclair and Coulthard's (1975) study of classroom discourse, prefiguring more recent calls for corpora to be included in language teacher education in relation to classroom discourse in the target language (Amador, Chambers & O'Riordan, 2006). It was not until the 1980s, however, that researchers began to emphasise that corpora could influence language teaching and learning in a number of fundamental ways. Carter and McCarthy (1988: Ch.3) stressed the importance of ensuring that the language taught corresponded closely to actual language use, in parallel with the emerging role of corpora as the basis for dictionaries and grammars. Johns (1986) and Tribble (1990) initiated reflection on the use of corpus data in the classroom.

The 1990s saw the publication of the first quantitative studies (Stevens, 1991; Cobb, 1997) of the effectiveness of using concordances as opposed to more traditional language-learning materials. More numerous but still limited in number are qualitative studies (Bernardini, 2002; Kennedy & Miceli, 2002; O'Sullivan & Chambers, 2006) which, crucially, include reports on learners' reactions to the consultation and analysis of native speaker corpus data. In the 1990s Granger (1998) and others made the case for learner corpora to be included in the data which are considered relevant for language learning, and this research also now influences the content of dictionaries and grammars. More recently multimodal corpora add new dimensions to the study of transcriptions of the spoken word, extending the scope of corpus linguistics to include non-verbal communication in a more detailed way than annotation conventions had previously allowed. As the list of references to this editorial illustrates, this research is published in two contexts, chapters in books including selected papers from conferences specialising in corpora and language learning, and articles in international journals on various aspects of second language acquisition.

This special issue contributes to the second of these contexts, arguably the more important as it has greater potential to bring corpus consultation and analysis into the mainstream of research and practice in language learning and teaching. In "Integrating learner corpora and natural language processing: A crucial step towards reconciling technological sophistication and pedagogical effectiveness", Sylviane Granger, Olivier Kraif, Claude Ponton, Georges Antoniadis and Virgine Zampa argue that, enriched with POS-tagging, learner corpora can provide a resource of interest to teachers and learners as well as researchers.

In research on the use of corpora in the language classroom, the theoretical and pedagogical underpinning tends to focus almost entirely on authenticity and learner autonomy. "Enhancing a process-oriented approach to literacy and language learning: The role of corpus consultation literacy" by Íde O'Sullivan takes a new approach, situating learning from corpus data in the context of research on the importance of process in language learning. "Concgramming: A computer-driven approach to learning the phraseology of English", by Chris Greaves and Martin Warren, complements the preceding article. They show how a more sophisticated version of concordancing software can provide the learner with data focusing on varied lexico-grammatical patterns rather than on the search word, in other words how data provided by the computer can assist the learner in the process of learning and of developing autonomy.

Two articles provide a quantitative and a qualitative study of the application of corpora by language teachers. Sabine Braun's study, "Integrating corpus work into secondary education: From data-driven learning to needs-driven corpora", is a welcome addition to the limited number of quantitative studies which have been published to date, and a rare example of a study of the use of corpora in secondary education. In "Learning English grammar with a corpus: Experimenting with concordancing in a university grammar course", Maria Estling Vannestål and Hans Lindquist illustrate the importance of teacher reflection in action research, as they report on how they altered their study in response to their learners' reactions. The authors of these two articles most certainly cannot be accused of painting a facile and overly optimistic picture of the inclusion of corpus data in language teaching. They report in detail on the obstacles which they encountered and the ways in which they reacted to them, thus providing valuable guidance to teachers and researchers wishing to follow in their footsteps. Finally "Using a multimodal concordancing tool for materials development", by Katherine Ackerley and Francesca Coccetta, is original in two ways. Firstly it shows how corpus data can be of use for learners at beginners' level, thus contradicting a widely held assumption. Secondly it uses a multimodal corpus to include non-verbal information in language-learning materials.

While these studies make an important contribution to the integration of corpora in the language learning and teaching environment, they represent only part of the potential of this developing area. Other domains not included here are parallel corpora, and corpora in areas such as translation studies, the teaching of literature, and language teacher education. These six articles do, however, make an important contribution to research on corpora and language learning in general, as they include not only studies of theoretical perspectives and technological advances, but also detailed quantitative and qualitative investigations of actual attempts to integrate corpora in language learning and teaching.

References

Amador, C. P., Chambers, A. and O'Riordan, S. (2006) Integrating a corpus of classroom discourse in language teacher education: The case of discourse markers. *ReCALL*, **18**(1): 83–104.

Bernardini, S. (2002) Exploring new directions for discovery learning. In: Kettemann, B. and Marko, G. (eds.) *Teaching and Learning by Doing Corpus Analysis*. Amsterdam: Rodopi, 165-182.

Carter, R. A. and McCarthy, M. J. (1988) *Vocabulary and Language Teaching*. Harlow: Longman.

Cobb, T. (1997) Is there any measurable learning from hands on concordancing? *System*, **25**(3): 301–315.

Granger, S. (ed.) (1998) *Learner English on Computer*. London & New York: Addison Wesley Longman.

Johns, T. (1986) Micro-concord: A language learner's research tool. *System*, **14**(2): 151–162.

Kennedy, C. and Miceli, T. (2002) The CWIC project: Developing and using a corpus for intermediate Italian students. In: Kettemann, B. and Marko, G. (eds.), *Teaching and Learning by Doing Corpus Analysis*. Amsterdam: Rodopi, 183–192.

McEnery, T. and Wilson, A. (1997) Teaching and language corpora. *ReCALL*, **9**(1): 5–14.

O'Sullivan, Í. and Chambers, A. (2006) Learners' writing skills in French: Corpus consultation and learner evaluation. *Journal of Second Language Writing*, **15**(1): 49–68.

Sinclair, J. and Coulthard, M. (1975) *Towards an Analysis of Discourse: The English Used by Teachers and Pupils*. Oxford: Oxford University Press.

Stevens, V. (1991) Concordance-based vocabulary exercises: a viable alternative to gap-fillers. In: Johns, T. and King, P. (eds.), *Classroom Concordancing: English Language Research Journal 4*. Centre for English Language Studies: University of Birmingham, 47–63.

Tribble, C and Jones, G. (1990) *Concordances in the Classroom. A Resource Book for Teachers*. Harlow: Longman.

Additional Reviewers