

## Wanted: Large Corpus, Simple Software. No Timewasters

Alex BOULTON  
ATILF, Equipe Crapel  
CNRS, University of Lorraine  
Nancy, France  
alex.boulton@univ-lorraine.fr

Data-driven learning involves the use of dedicated concordancers to explore large language corpora. Or does it? Willis (1998) reported on the use of manual concordancing involving nothing more than texts, learners and a blackboard, a technique which Johns (1993) recognised as DDL, and similar activities can be found described as DDL today (e.g. Tyne, in press). The appeal of such activities is of course that they reduce or eliminate the need for technology altogether. This is certainly an advantage, as the technology itself is frequently cited as a major obstacle, and possibly the prime reason why DDL has not become more mainstream practice (e.g. Boulton, 2010). The problem with this approach is that it does not make use of the potential of computers to help in the inductive, discovery-based noticing process. An alternative path is to simplify the technology as much as possible – either the corpus itself or the associated software. Firstly, there has been some reaction against the early corpora (large, general-purpose, linguistically oriented) in favour of small, *ad hoc* corpora of a specific language genre (e.g. Ghadessy *et al.*, 2001), often designed with explicitly pedagogical aims in mind (Braun, 2007). Similarly, great strides have been made in making concordancers more user-friendly, but technology still remains a difficulty cited in many contemporary papers (e.g. Rodgers *et al.*, 2011). What is it about the concepts of ‘corpus’ and ‘concordancer’ that keeps them out of mainstream classroom practice?

A third way is to begin not with corpus or concordancer, but with the learners and what they already do – in other words, to bring our work closer to them rather than making them come to us. In what ordinary, everyday activities are learners involved in using computers to search for information outside the language classroom? Most obviously, in browsing the Internet. A small number of studies attempt to show how this can be used for language learning (e.g. Todd, 2001; Chinnery, 2008; Sha, 2010); the temptation then is to wonder whether the Internet might not serve as a substitute corpus, and Google as a substitute concordancer in a more-or-less DDL approach. This might seem inappropriately iconoclastic at a TaLC conference, but DDL and corpus linguistics have broken more than one ideological barrier themselves. If these resources had been available in the 1980s, we would likely be working in a very different field today.

This paper sets out to see whether such an approach can be considered a form of DDL, at least in abstract terms, then goes on to test the potential and limits of the Internet and Google for linguistic searches relevant to language learners. In conclusion, it is argued that the processes are in fact not entirely dissimilar, and that they can provide one way in to DDL – immediately useful for all, and potentially leading some on to more prototypical DDL activities with corpus and concordancer.

### References

- Boulton, A. (2010). 'Data-driven learning: Taking the computer out of the equation.' *Language Learning* 60: 534-572.
- Braun, S. (2007). 'Integrating corpus work into secondary education: From data-driven learning to needs-driven corpora.' *ReCALL* 19: 307-328.
- Chinnery, G. (2008). 'You've got some GALL: Google-assisted language learning.' *Language Learning & Technology* 12: 3-11.
- Ghadessy, M., Henry, A. & Roseberry, R. (eds). (2001). *Small Corpus Studies and ELT: Theory and Practice*. Amsterdam: John Benjamins.
- Johns, T. (1993). 'Data-driven learning: An update'. *TELL&CALL* 2: 4-10.
- Rodgers, O., Chambers, A. & LeBaron, F. (2011). Corpora in the LSP classroom: A learner-centred corpus of French for biotechnologists. *International Journal of Corpus Linguistics* 16: 392-358.
- Sha, G. (2010). 'Using Google as a super corpus to drive written language learning: A comparison with the British National Corpus.' *Computer Assisted Language Learning* 23: 377-393.
- Todd, R. (2001). 'Induction from self-selected concordances and self-correction.' *System* 29: 91-102.
- Tyne, H. (in press). 'Corpus work with ordinary teachers: Data-driven learning activities.' In Thomas, J. & Boulton, A. (eds). *Input, Process and Product: Developments in Teaching and Language Corpora*. Brno: Masaryk University Press.
- Willis, J. (1998). 'Concordances in the classroom without a computer.' In Tomlinson, B. (ed). *Materials Development in Language Teaching*. Cambridge: Cambridge University Press, 44-66.