

Rigour, Relevance, Reflection: CL Methodology Through a Critical Lens

Gerlinde Mautner (Jenna University of Economics and Business, Austria)

The aim of this paper is to critically examine the combination of corpus linguistics (CL) and Critical Discourse Studies (CDS). It will do so from two angles: (i) an epistemological angle, discussing questions of knowledge generation and (ii) a social and interdisciplinary angle, debating the different questions, mindsets and skills that different kinds of researchers bring to the task, as well as the potential, often underexplored, for interdisciplinary cooperation.

The epistemological angle

Although the approach combining CL and CDS is now well established, with a twenty-year history under its belt (Hardt-Mautner, 1995; Baker, 2006), it still requires careful and indeed ongoing epistemological scrutiny. For it is all too easy to be lured into a false sense of security about the nature of the evidence merely because the analysis is computer-based. To avoid rash interpretative shortcuts, the following observations ought to be borne in mind (see Mautner, 2016).

(1) Evidence that corpus-linguistic software produces is rarely self-explanatory. The program may be able to identify patterns, but it falls to the analyst to spot interesting outliers, and relate both the expected and the unexpected to the context in which the texts concerned were produced and received.

(2) The presence of certain linguistic forms in a text does not lead straight to its social meaning, function and significance, and we should not pretend that it does. Although this is a truism barely worth reiterating, it is still quite common for researchers to read too much into their corpus evidence. There is a certain irony in this, because 'overinterpretation' is one of the fallacies that the use of CL is supposed to help guard against (O'Halloran & Coffin, 2004).

(3) However sophisticated, every analytical tool is very good at some things but not so good at others. Dutifully acknowledging a tool's limitations is not only a matter of scholarly integrity, but also an important step towards finding ways in which complementary tools may help compensate for any deficiencies.

(4) Every corpus is an artefact constructed for a particular purpose, and caution is required when generalizing from it. As McEnery and Hardie point out (2012, p. 26), 'we must not confuse corpus data with language itself'.

(5) Furthermore, there is a tendency for computer-based evidence to be rated more highly than data gathered and processed manually. This judgement may be justified in certain circumstances but misguided in others. It is most certainly counterproductive if one aims at triangulation and a rich mixed-methods approach.

(6) Quantitative statements about the linguistic behaviour of an item only make sense in comparison with its use in a different corpus.

(7) Both quantitative and qualitative statements about the data have their uses, and the analysis will be all the richer if both can be made. But the boundaries between them should be clear at all times, and vague quantifiers (e.g., *a few, some, many, a lot*) avoided.

The social and interdisciplinary angle

If epistemological critique helps us understand how knowledge is generated, it is equally worthwhile to reflect on who the key actors are in this process, and where they are positioned in the disciplinary landscape. There are two perspectives that can be usefully applied, within linguistics and outside it.

Within linguistics, the chasm appears to be widening between those researchers who are able and willing to develop new software and those who are not. The latter group is keen to have powerful analytical tools, but the reason why they want them is because they are tackling a research question at the interface of discourse and society. On the other hand, those who concentrate on the computational side of things often seem oblivious to the socio-political implications of their work. If there are calls for discourse analysts to learn how to write code and more generally to improve the rigour of their research, there should at least be corresponding calls for computational linguists to focus more on relevance, and engage more deeply in the social functions of discourse.

Finally, we should not forget that there is a large and prolific community of discourse scholars outside linguistics, in areas such as organizational theory and behaviour for example. In those communities, corpus-based approaches are still virtually unknown. They use software for qualitative data analysis, such as NVivo or ATLAS.ti, rather than concordancers. Their hero is Foucault, not Sinclair, and they are generally concerned much more with discursive macro-structures than with linguistic detail. It should benefit both sides of the disciplinary divide to initiate a focused and productive conversation between them.

References

- Baker, P. (2006). *Using corpora in discourse analysis*. London and New York: Continuum.
- Hardt-Mautner, G. (1995). Only connect: Critical Discourse Analysis and corpus linguistics, *UCREL Technical Paper 6*. Lancaster: University of Lancaster. Available at http://www.comp.lancs.ac.uk/ucrel/tech_papers.html.
- McEnery, T. & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge: CUP.
- Mautner, G. (2007). Mining large corpora for social information: The case of *elderly*. *Language in Society* 36, 51–72.
- Mautner, G. (2009) Corpora and critical discourse analysis. In P. Baker (Ed.). *Contemporary corpus linguistics* (pp. 32–46). London: Continuum.
- Mautner, G. (2016). Checks and balances: How corpus linguistics can contribute to CDA. In R. Wodak & M. Meyer (Eds.). *Methods of critical discourse analysis*. (pp. 154–179). London, etc.: Sage.
- O'Halloran, K. and Coffin, C. (2004). Checking overinterpretation and under-interpretation. Help from corpora in critical linguistics. In A. Hewings, C. Coffin & K. O'Halloran (Eds.). *Applying English grammar* (pp. 275–297). London: Arnold.