

Abs-171

Rachelle Freake (Queen Mary, University of London)

Methodological challenges in bilingual corpus-assisted discourse analysis

This paper presents some of the challenges arising from the use of corpus-assisted discourse studies (CADS) of bilingual, non-parallel corpora (Baker et al., 2008; Partington, 2004; Qian, 2010). Since corpus linguistics and discourse analysis ultimately focus on “real language use” rather than theoretically constructed examples (e.g. Stubbs, 1996; Halliday & Matthiessen, 2004), it follows that data should be as varied as the population, and to an even larger extent when the population is ethnolinguistically diverse. Data for CADS research, then, can pose numerous methodological issues to researchers, particularly if they are using CADS in sociolinguistic research and drawing on multilingual, multicultural data.

Examples for this paper are drawn from two cases in Canada, a country that contains a diverse population indexed by two official languages, English and French (Gal and Irvine, 1995; Heller, 1999). The first bilingual corpus is comprised of texts submitted to the 2007 Bouchard Taylor Commission in Quebec (Freake et al., 2011). The second bilingual corpus consists of Canadian newspaper articles from 2009. In both cases, the unilingual components of the bilingual corpora are of unequal sizes and index a sector of Canadian society wherein a specific ethnolinguistic population dominates. In the first case, the majority group in Quebec is ethnolinguistically French-Canadian, and so it follows that the French component of the Bouchard Taylor corpus is comparatively larger. In the second instance, there are more English speakers (newspaper producers and readers) in Canada than French speakers, hence the larger English component of the bilingual newspaper corpus. These corpus differences, which arose from the use of naturally-occurring data in a particular but not unusual context, led to a number of challenges which were addressed in the research. The challenges highlighted here are the following:

Studying single-language corpora of texts produced by multiethnic, linguistically-indexed populations;

The essentialization of differences and the reification of pre-existing ethnolinguistic categories;

The semantic preference of “cognates” in corpora of different languages; and

The comparison of keywords derived from comparator corpora of different languages of different sizes.

Overall, CADS is found to be a valuable method and approach to data, providing unique top-down and bottom-up perspectives. However, when CADS is used to examine sociolinguistic data, it is not without potential shortfalls. To address these shortfalls, this paper presents some methodological solutions that were found to be useful in the Canadian context. These solutions may be useful to other researchers employing CADS in other multilingual, multicultural contexts.

References:

- Baker, P., Gabrielatos, C., Khosravinik, M., Krzyzanowski, M., McEnery, T. & Wodak, R. (2008). A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press. *Discourse & Society*, 19(3), 273-306.
- Freake, R., Gentil, G. & Sheyholislami, J. (2011). A bilingual corpus-assisted discourse study of the construction of nationhood and belonging in Quebec. *Discourse & Society*, 22(2), 21-47.
- Gal, S. & Irvine, J. T. (1995). The boundaries of languages and disciplines: How ideologies construct

- difference. *Social Research*, 62 (4), 967-1001.
- Halliday, M. A. K. and Matthiessen, C. (2004). *An Introduction to Functional Grammar*. London: Arnold.
- Heller, M. (1999a). Heated language in a cold climate. In J. Blommaert (Ed.), *Language ideological debates* (pp. 143-170). Berlin: Mouton de Gruyter.
- Partington, A. (2004). Corpora and discourse, a most congruous beast. In A. Partington, J. Morley & L. Haarman (Eds), *Corpora and Discourse* (pp. 11–20). Bern: Peter Lang.
- Qian, Y. (2010). *Discursive construction around terrorism in the People's Daily (China) and The Sun (UK) before and after 9.11*. Oxford: Peter Lang.
- Stubbs, M. (1996). *Text and corpus analysis: Computer-assisted studies of language and culture*. Oxford, UK & Cambridge, USA: Blackwell Publishers.