

Mapping Undergraduate Culture Using the Personal Blog: A Corpus-Based Approach

Vincent B Y Ooi,¹ Peter K W Tan¹
and Andy K L Chiang¹

Abstract

The exponential increase in the number of weblogs (or 'blog') on the Web in the past few years warrants a serious consideration of this new text type in the corpus linguistic community. In this paper, we examine the mapping of linguistic and cultural issues relating to undergraduate blogs in two communities using the WMatrix approach (cf Rayson 2005; Rayson and Garside 2000). WMatrix represents an integrated corpus linguistic software environment that not only allows the usual concordancing and frequency counts but also provides a semantic content analysis based on 21 categories (Piao et al 2005).

Huffaker and Calvert (2005) observe that blogs are a distinctive form of computer-mediated communication for a variety of reasons, including the means for users to connect their blogs to other blogs by adding links to them. Many undergraduates, for instance, nowadays express themselves online and are connected to one another to form an online/virtual community with their attendant cultural concerns and linguistic patterning. The personal blog written in English also tends to include computer-mediated communication features (Crystal 2003) that are said to impinge on the reformulation of English in the future (Graddol 2007). In addition, it would be interesting to examine the robustness (or otherwise) of WMatrix in handling newer varietal features in non-Anglo English contexts (Kachru 2005).

This paper will therefore analyse 2 corpora, with a sampling size of 100000 words for each corpus, from the U.K. and Singapore respectively (each including an equal number of male and female blogs). Using WMatrix, the paper proposes to examine how male undergraduates differ from their female counterparts in their range of topics/interests and lexico-grammatical choices, and the extent to which undergraduates from these two different communities do share common interests.

References

- Crystal, D. 2003. Electronic variation, in Cambridge Encyclopedia of the English Language 2nd edition, 423–33.
Graddol, D. (ed). 2007. *Changing English*. London: Routledge.

¹ National University of Singapore
e-mail: vinceooi@nus.edu.sg, petertan@nus.edu.sg, stackl@nus.edu.sg

- Huffaker, D.A., and S.L. Calvert. 2005. Gender, identity, and language use in teenage blogs. *Journal of Computer-Mediated Communication*, 10(2): 1 [currently at <http://jcmc.indiana.edu/vol10/issue2/huffaker.html>]
- Kachru, B.B. 2005. *Asian Englishes: beyond the Canon*. Hong Kong: Hong Kong University Press.
- Piao, S, D. Archer, O. Mudraya, P. Rayson, R. Garside, T. McEnery and A. Wilson. 2005. A large semantic lexicon for corpus annotation, in *Proceedings from the Birmingham Corpus Linguistics Series* July 14–17, 2005. [currently at http://www.corpus.bham.ac.uk/PCLC/cl05_est_lex_paper_final.doc]
- Rayson, P. 2005. Wmatrix: a web-based corpus processing environment, Computing Department, Lancaster University. [currently at <http://www.comp.lancs.ac.uk/ucrel/wmatrix>]
- Rayson, P. and R. Garside. 2000. Comparing corpora using frequency profiling, in *ACL Proceedings of the Workshop on Comparing Corpora*, Hong Kong. [currently at http://www.comp.lancs.ac.uk/computing/users/paul/publications/rg_ac12000.pdf]