*I don't know, I don't think,* and *I don't want:* an exploratory, corpus-based study of formulaic language in American blogs

Federica Barbieri
Swansea University

Formulaic language has been studied from different theoretical perspectives and under different rubrics. Over the past decade or so, within text analysis, adopting corpus linguistics methods, many studies have investigated 'lexical bundles' (a.k.a., n-grams or clusters). Lexical bundles are recurrent sequences of words in a text. They are not necessarily syntactically complete and typically not idiomatic, but they are important building blocks of discourse in spoken and written registers, allowing researchers to uncover unexpected structural and lexico-grammatical characteristics of texts (Biber et al., 2004; Biber & Barbieri, 2007; Csomay, 2013). For example, Biber et al. (2004) found that US classroom teaching has far more lexical bundle types than casual conversation, and over one third of these are stance-related bundles.

Computer-mediated discourse, including blogging and micro-blogging, has been the focus of increasing interest within discourse analysis. However, to date, relatively little is known about the linguistic characteristics of web-based registers or genres (Grieve et al., 2010; Titak & Robertson, 2013), including formulaic language. Accordingly, the present study explores lexical bundles in blogs.

The study investigates lexical bundles in a 2.2 million-word corpus of personal blogs in American English (2003-2005), collected following geographic and demographic criteria (Grieve et al., 2010). Preliminary findings reveal that there are over 80 different (i.e., types) 4-5 word lexical bundles occurring at least 40 times per million words in blogs. Thus, blogs are similar to classroom teaching in frequency of bundle types, and have far more different bundles than casual conversation (cfr. data in Biber et al., 2004). Further, a preliminary functional classification of bundles reveals that blogs' lexical bundles are overwhelmingly 'stance expressions'. Findings of structural and functional analyses of lexical bundles in American blogs will be compared with previous findings on lexical bundles in spoken and written registers, to situate blogs on the 'oral-literate continuum'.

[300 words]

References:

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes, 26*, 263-286.

Biber, D., Conrad, S., & Cortes, V. (2004). *If you look at...*: Lexical bundles in university teaching and textbooks. *Applied Linguistics, 25*(3), 371-405.

Csomay, E. (2013). Lexical bundles in discourse structure: A corpus-based study of classroom discourse. *Applied Linguistics, 34*(3), 369-388.

Grieve, J., Biber, D., Friginal, E., & Nekrasova, T. (2010). Variation among blogs: A multi-dimensional analysis. In A. Mehler, S. Sharoff & M. Santini (Eds.), *Genres on the web: Computational models and empirical studies: Text, speech and language technology* (pp. 45-71). Dordrecht: Springer.

Titak, A., & Roberson, A. (2013). Dimensions of web registers: an exploratory multi-dimensional comparison. *Corpora, 8*(2), 235-260.