

## The grammar and discourse of American personal blogs: A lexical bundles approach.

Federica Barbieri (Swansea University)

Research on the linguistic characteristics of internet genres or registers is in its infancy. Blogs, however, are one of the oldest genres of the internet, and, along with wikis, Facebook and Twitter posts, internet forums, and chat, one of the most perceptually salient to users (Biber et al. 2015). Yet, while multi-dimensional analyses of internet corpora have begun to paint a picture of the linguistic make-up of web-based registers (Titak & Roberson 2013; Biber & Egbert 2016), including blogs (Grieve et al. 2010; Hardy & Friginal 2012), we know relatively little about the lexico-grammatical characteristics of even these early internet registers.

Lexical bundles are frequently occurring sequences of words identified automatically (Biber et al., 1999); they tend to be non-idiomatic and structurally incomplete. Yet, they are important building blocks of discourse, and can reveal unexpected structural and lexico-grammatical characteristics of texts (Biber et al., 2004). Accordingly, taking a corpus-driven approach, the present study aims to address the above described research gap by looking at lexical bundles in blogs. Specifically, the study investigates the discourse function, person reference, and the grammatical structure of lexical bundles, in order to shed light on the structural and discourse characteristics of American blog writing.

The study is based on a 2.2 million-word corpus of blogs in American English (2003-2005; Grieve et al., 2010), comprising 500 blogs from different US states and different social groups. Exploratory analysis of the frequency of lexical bundles reveals that bundles in personal blogs are overall infrequent – more infrequent than in academic prose and conversation.

Analysis of discourse function applies the functional taxonomy in Biber et al. (2004) and reveals that bundles in blogs are overwhelmingly ‘stance’ bundles, accounting nearly half of all bundles; ‘referential’ bundles follow, while ‘discourse organizing’ bundles are negligible. Finally, a sizeable group of bundles not fitting the three established categories above were classified as ‘narrative’ bundles.

Analysis of grammatical person reference reveals that bundles with no reference dominate, but bundles with first person reference follow with nearly 40%.

Finally, exploratory analysis of grammatical structure shows that verb phrase (VP) based bundles represent over 60% of all bundles, while noun-phrase (NP) and prepositional phrase (PP) based bundles are far less frequent, at ca. 20% each. The present study will also report on a more detailed analysis of grammatical structure, applying Biber et al.’s (1999) structural taxonomy.

Taken together, the findings from the present study reveal that American personal blogs are characterized by a unique combination of discourse and lexico-grammatical features, which reflect the synergy of mode (writing) and communicative purpose. For example, like conversational bundles, bundles in blogs tend to be VP-based; however, blog writing relies on NP- and PP-based bundles more than conversation. Like oral registers, blogs rely heavily on stance bundles; however, like literate registers, they also use referential expressions extensively.

### References

- Biber, Douglas, Conrad, Susan, & Cortes, Viviana. 2004. If you look at...: Lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25(3), 371-405.
- Biber, Douglas, & Egbert, Jesse. 2016. Register variation on the searchable web: A multidimensional analysis. *Journal of English Linguistics*, 44(2), 95-137.
- Biber, Douglas, Egbert, Jesse, & Davies, Mark. 2015. Exploring the composition of the searchable web: a corpus-based taxonomy of web registers. *Corpora*, 10(1), 11-45.
- Biber, Douglas, Johansson, Stig, Leech, Geoffrey, Conrad, Susan, & Finegan, Edward. 1999. *The Longman Grammar of Spoken and Written English*. London: Longman.
- Grieve, Jack, Biber, Douglas, Friginal, Eric, & Nekrasova, Tatiana. 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.

- Hardy, Jack A., and Eric Friginal. 2012. "Filipino and American online communication and linguistic variation." *World Englishes* 31 (2):143-161.
- Titak, Ashley, & Roberson, Audrey. 2013. Dimensions of web registers: an exploratory multidimensional comparison. *Corpora*, 8(2), 235-260.