

*I don't know why...*: lexical bundles in American blogs. A grammar-in-discourse approach.

Blogs 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). However, while multi-dimensional analyses of internet language corpora have begun to outline 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 (e.g., *I don't know what, the rest of the*) are frequently occurring sequences of words identified automatically (Biber et al., 1999). They are usually structurally incomplete and typically non-idiomatic, but they function as important building blocks of discourse, and have been shown to encapsulate unexpected structural and lexico-grammatical characteristics of texts (Biber et al., 2004). Accordingly, taking a corpus-driven approach, Author (2016) investigated the discourse function of lexical bundles in American blogs, applying Biber et al.'s (2004) taxonomy. The study revealed that bundles in blogs tend to be 'stance' bundles. Stance bundles indeed account for nearly half of all bundles; 'referential' bundles follow, while 'discourse organizing' bundles are negligible. The study further showed that blog writers also rely on bundles that primarily support a narrative function and thus do not fit the three main categories in Biber et al.. Taken together, these findings suggest 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 defining.

The present study extends this work by investigating the grammatical structure of lexical bundles in 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 analyses of the grammatical structure show 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 study investigates the mapping of grammatical structures on communicative functions, showing, for example, that narrative expressions tend to be VP-based and more fragmented than stance bundles. Additionally, the study will report on a more fine-grained analysis of grammatical structure based on Biber et al.'s (1999) taxonomy of structural patterns.

Select 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 multi-dimensional analysis. *Journal of English Linguistics*, 44(2), 95-137.
- 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.
- Titak, Ashley, & Roberson, Audrey. 2013. Dimensions of web registers: an exploratory multi-dimensional comparison. *Corpora*, 8(2), 235-260.