Lexical Bundles in Academic Discourse

Nichole M. Oberheu

Colorado State University
Abstract

Lexical bundles are important to the fluid production of English; subsequently, it is important that both English language learners and English language teachers understand the many forms and functions of lexical bundles. The use of lexical bundles differs depending on register, mode of communication, and formality of context. Even in the academic register, lexical bundle use varies across disciplines. The difference of lexical bundle use between spoken and written academic discourse, and across academic registers is examined. Additionally, the extent to which a generalized list of academic lexical bundles can be formulated for pedagogical purposes is discussed.

Keywords: academic English, fluency, lexical bundles, register
Decades of phraseological research have demonstrated the importance of multi-word collocations, or lexical bundles, in the fluid production of English (Biber, Johansson, Conrad, & Finegan, 1999). Lexical bundles comprise approximately 80% of natural English (Hyland, 2008). Biber (2009) notes that whereas other languages use a broad range of inflectional morphology to provide additional meaning to communication, there is a relatively limited range of inflectional morphology in English; rather, English utilizes grammatically functional words, such as lexical bundles, to supplement additional meaning in discourse. This indicates that it is critical for English language learners to acquire knowledge of lexical bundles to establish fluency, not only for casual conversation but also in formal, academic communication.

Lexical bundles differ across registers regarding the form and purpose of communication (Biber et al., 1999). The function of lexical bundles is dependent on the reason and formality of communication, and whether communication is made via oral or written discourse. As Biber et al. (1999) points out, “most lexical bundles in conversation are building blocks for verbal and clausal structural units, while most lexical bundles in academic prose are building blocks for extended noun phrase or prepositional phrases” (p. 992). Therefore, conversational lexical bundles can be seen as primarily clausal in nature, and lexical bundles in academic prose are primarily phrasal (Biber et al., 1999). Recently, there has been a considerable amount of research pertaining to the functional variations of lexical bundles within academic discourse; particularly, how lexical bundle use differs across academic disciplines and between spoken and written academic discourse. In the following paper I will provide a definition of lexical bundles, the general
LEXICAL BUNDLES IN ACADEMIC DISCOURSE

characteristics of lexical bundles in academic discourse and how the functionality of bundles varies by communicational form and purpose. I will conclude with describing new research in overcoming the inherent differences of lexical bundle function in academic discourse, and the potential of creating a generalized list of academic bundles for pedagogical use.

Definition and Description

According to Biber et al. (1999) lexical bundles are strings of three or more words that frequently occur together. These strings of words need to occur at least ten times per million in a corpus to be considered a lexical bundle. Lexical bundles are semantically transparent; that is, they are non-idiomatic. A single word cannot be substituted for the lexical bundle and longer lexical bundles are created by adding to or combining shorter lexical bundles. They are not fixed structural units; in fact, lexical bundles tend to extend across structural units.

Biber and Barbieri (2007) describe three major functional categories of lexical bundles: Stance expressions, discourse organizers, and referential expressions. Stance expressions describe assessments of certainty of a particular phrase or clause and are divided into five subcategories: Epistemic (for example, I know that), desire (for example, I want you to), obligation (for example, you must do), intention/prediction (for example, what you will do), and ability (for example, they will be able). According to these researchers, discourse organizing bundles are used to bridge past and future ideas and are divided into three subcategories: topic introduction (for example, what we will do), topic elaboration/clarification (for example, has to do with the), identification/focus bundles (for example, those of you who). The last major functional category of lexical
bundles is referential expression. This type of bundle references a particular attribute, idea, or object and can be sub-categorized into three fields: imprecision indicators (for example, *something like that*), specification of attributes (for example, *a little bit of*), and time/place/text reference (for example, *as shown in figure*) (Biber & Barbieri, 2007).

**Lexical Bundles and Written Vs. Spoken Registers**

Biber (1999) cites that lexical bundles differ depending on register. As discussed above, lexical bundles in spoken, conversational discourse are functionally different than those that occur in written registers (Biber et al., 1999). Conversational bundles tend to be comprised mainly of stance bundles (Biber, Conrad & Cortes, 2004). This is due to the fact that a conversation is comprised of two or more individuals creating and negotiating meaning in real time (Biber, Conrad & Cortes, 2004). Written bundles, such as academic prose, are generally comprised of informational communication and therefore tend to use discourse organizing and referential bundles more frequently (Biber & Barbieri, 2007).

Lexical bundle patterns vary in regard to the form and purpose of communication (Biber et al., 1999). Research has been conducted to discover the distinction between written academic bundles and spoken academic bundles (Biber & Barbier, 2007; Biber, Conrad & Cortes, 2004). Spoken, or oral, academic bundles, for the purposes of this paper, are described as those found in classroom or lecture talk. Similarly, I categorize written academic bundles as those found in textbooks and academic prose.

Oral academic bundles differ from both typical conversational bundles and typical bundles found in academic prose (Csomay & Cortes, 2010). Oral academic bundles are more of a hybrid between typical conversational bundles and more formal written academic bundles. As part of the oral register, academic lectures utilize stance bundles to
influence audience perceptions similar to those used in conversation; however, unlike conversations, lectures are created in advance, contain a large quantity of information that requires the listener to make correlations, and are conducted in the style of a monologue. Subsequently, referential expressions and discourse organizing bundles are adopted to guide the audience in associating visual props to verbal topics and to “help listeners predict what's coming next” (Nesi & Basturkmen, 2006, p. 301). As Biber and Barbieri (2007) describe, discourse organizers and referential bundles in teaching serve two purposes: topic introduction/focus and identification/focus bundles. Topic introduction/focus bundles suggest that a new subject is beginning by employing verbs of intention or desire with first and/or second person pronouns (such as, *I want to talk about*) or by illuminating visual aids that tie into the lecture topic (such as, *if you look at*). Identification/focus bundles are used to preview or summarize an important point (such as, *one of the things*).

According to Biber et al. (1999), the majority of lexical bundles in the written academic register are referential. Biber and Barbieri (2007, p. 282) state that written academic bundles, “regardless of purpose, are highly 'informational', 'elaborated in reference', and marked for 'impersonal' styles;” these attributes necessitate frequent use of referential bundles. Referential bundles are mostly incomplete structural units that typically span across a noun phrase and the beginning of a prepositional phrase (Biber et al., 1999). When a lexical bundle is complete in academic writing it usually denotes a discourse organizing expression (Biber et al., 1999). Discourse organizing expressions are the next most frequently applied lexical bundle unit in written academic discourse. Just as in academic lectures, referential and discourse organizing bundles are used to
signal important points, elicit connections between topics or ideas, and preview and summarize information.

**Lexical Bundles and Academic Discipline**

In addition to the distinction between bundle functions in written versus spoken academic discourse, there is also significant variation of lexical bundle use across academic disciplines (Hyland, 2008). In a corpus-driven study of texts within the physical and social sciences, Hyland (2008) discovered that the function of lexical bundles differed dramatically between disciplines. Through an adaptation of Biber and Barbieri's (2007) lexical bundle taxonomy, Hyland (2008) describes three broad features to categorize the types of lexical bundles discovered in his study: research-oriented bundles, text-oriented bundles, and participant-oriented bundles. Research-oriented bundles assist writers in organizing empirical information. Text-oriented bundles structure arguments, link ideas, or make references to previously presented information. Participant-oriented bundles demonstrate authorial opinion or engage readers with the text (Hyland, 2008).

Hyland (2008) found that the physical sciences utilized research-oriented bundles, whereas text-oriented and participant-oriented bundles were employed with more frequency in the social sciences. Hyland (2008) explained that this distinction is due to the ideological nature of each discipline. The physical sciences value empirical information over interpretive analysis and tend to minimize the role of the researcher and let the research speak for itself (Hyland, 2008). Participant-oriented bundles were utilized in the physical sciences to a small degree to guide audience through empirical results and to ensure that interpretations are acknowledged without ambiguity (Hyland, 2008).

In contrast, the rhetorical style of the social sciences is much more interpretive and
persuasive. The text-oriented bundles allow authors to direct readers through the discourse by highlighting arguments and goals, and specifying particular circumstances in which an assertion can be considered correct (Hyland, 2008). Participant-oriented bundles are used within the social sciences to communicate uncertainty through hedges and anticipatory-it structures (such as, *may be due to* and *it is possible that*) therefore protecting the author from statements that could be conceived as false (Hyland, 2008).

**Towards a List of Academic Formulas**

As demonstrated above, lexical bundles have a range of functions throughout academic registers and disciplines. With this variation, some scholars refute the notion that a generalized compilation of academic lexical bundles can be formulated for pedagogical use (Hyland, 2008). However, in a recent study, Simpson-Vlach and Ellis (2010) apply different methods to identify pertinent lexical bundles in academic discourse. In past studies, researches have focused almost exclusively on statistical frequency to identify lexical bundles in a particular corpus. This procedure has resulted in long lists of lexical bundles that are nearly impossible to consolidate for pedagogical purposes; therefore, Simpson-Vlach and Ellis (2010) decided to adopt criteria that combine quantitative and qualitative methods. The authors argue that applying a mixed methods approach will elicit more meaningful results than frequency analysis alone. As Simpson-Vlach and Ellis (2010) explain, “relying solely on frequency means that some distinctively useful but lower frequency phrases whose component words are highly unlikely to occur together by chance will not make it to the top of the frequency-ordered list.” In addition to frequency identification, Simpson-Vlach and Ellis (2010) use a statistical method called mutual information (MI) to measure cohesiveness, and brought
in experts to validate and prioritize the most useful bundles. The researchers also used a lower cut off frequency and included three-word bundles (Simpson-Vlach & Ellis, 2010). It was argued that by “casting a wider net” they would be able to find more bundles that transcend disciplines; after the lexical bundles were tabulated, the bundles were then analyzed for relevance using MI and expert discrimination (Simpson-Vlach & Ellis, 2010, p. 509). The corpus that they drew upon was comprised of both written and spoken discourse, and spanned across six academic disciplines (Simpson-Vlach & Ellis, 2010). After identifying the most relevant and useful bundles across disciplines, the authors formulated the Academic Formulas List (Simpson-Vlach & Ellis, 2010). Simpson-Vlach and Ellis (2010) believe that this compilation is representative of all disciplines surveyed in both written and spoken academic registers. This study indicates that, contrary to Hyland's (2008) assertion, there are bundles that exist in all registers despite particular rhetorical device variations.

The study also highlights the potential of new research methods to identify lexical bundles in academic corpora. While Biber (2009) notes that MI is not a satisfactory measure to discover lexical bundles because it identifies collocations regardless of word order, Simpson-Vlach and Ellis (2010) also utilize expert informants to recognize salient bundle forms. In doing so, the researchers were able to create a concise list of bundles that are applicable to a range of academic disciplines and communication forms.

Conclusion

Lexical bundles are important in the creation of fluid academic discourse; whether communication occurs in the form of a written prose or classroom lecture, lexical bundles assist in structuring the textual meaning. Lexical bundles are defined as strings of
uninterrupted words occurring frequently throughout a corpus. They are not idiomatic, nor are they typically considered complete structural units; instead, they span units (Biber et al., 1999). Lexical bundle functions vary depending on the form and purpose of communication (Biber & Barbieri, 2007). Lexical bundle functions in academic discourse vary between spoken and written contexts and across academic disciplines (Biber, Conrad, & Cortes, 2004; Biber & Barbieri, 2007; Hyland, 2008). Most researchers utilized methodologies that focused on statistical frequency alone to allocate bundles; recently, Simpson-Vlach and Ellis (2010) have devised a novel strategy in assessing lexical bundles in academic discourse by combining qualitative and quantitative methodologies. This research resulted in the creation of the Academic Formula's List, a generalized list of academic bundles that spans disciplines and registers. The Academic Formula’s List may prove pedagogically potent in the English language classroom.
References


doi:10.1016/j.esp.2006.08.003


doi:10.1093/applin/25.3.371


doi:10.1093/applin/amp058