



# Lexical bundles in university spoken and written registers

Douglas Biber \*, Federica Barbieri

*English Department, Northern Arizona University, Flagstaff, AZ 86011-6032, United States*

---

## Abstract

Lexical bundles – recurrent sequences of words – are important building blocks of discourse in spoken and written registers. Previous research has shown that lexical bundles are especially prevalent in university classroom teaching, where they serve three major discourse functions: stance expressions, discourse organizers, and referential expressions.

The present study extends this line of research by investigating the use of lexical bundles in a wide range of spoken and written university registers, including both instructional registers and student advising/management registers (e.g., office hours, class management talk, written syllabi, etc.). The findings show that lexical bundles are even more prevalent in non-academic university registers than they are in the core instructional registers. Most surprisingly, lexical bundles are very common in written course management (e.g., course syllabi), in contrast to previous research which showed bundles to be much more common in speech than in writing.

© 2006 The American University. Published by Elsevier Ltd. All rights reserved.

---

## 1. Introduction

There have been many studies of written academic discourse over the past 20 years, focusing mainly on characteristic lexico-grammatical features or on ‘genre’ characteristics. The journal *English for Specific Purposes* has been one of the major forums for descriptive research of this type, while more recently, the *Journal of English for Academic Purposes* has

---

\* Corresponding author.

E-mail address: [douglas.biber@nau.edu](mailto:douglas.biber@nau.edu) (D. Biber).

focused on academic varieties and the pedagogical implications of descriptive studies. Recent book-length treatments of this type include Atkinson (1996), Hyland (1998), Swales (2004), and the edited collections in Flowerdew (2002), Hewings (2001), and Markkanen and Schroder (1997).

Until recently, there have been many fewer studies of spoken academic discourse. However, the availability of corpora such as the Michigan Corpus of Academic Spoken English (MICASE) (see Powell & Simpson, 2001; Simpson, Briggs, Ovens, & Swales, 2002) has resulted in recent investigations of the typical linguistic features found in spoken university language (see, e.g., Lindemann & Mauranen, 2001; Poos & Simpson, 2002).

One especially productive approach to the study of spoken academic discourse has been the description of longer lexical phrases, chunks, and idioms (see, e.g., DeCarrico & Nattinger, 1988; Khuwaileh, 1999; Nattinger & DeCarrico, 1992; Schmitt, 2004; Simpson, 2004; Simpson & Mendis, 2003). These studies are part of a growing research tradition focusing on the use of multi-word pre-fabricated expressions in English (see the reviews in Ellis, 1996; Howarth, 1996; Van Lancker-Sidtis & Rallon, 2004; Weinert, 1995; Wray, 2002; Wray & Perkins, 2000).

Multi-word sequences have been studied under many rubrics, including ‘lexical phrases’, ‘formulas’, ‘routines’, ‘fixed expressions’, and ‘pre-fabricated patterns’ (or ‘pre-fabs’). These approaches all define the object of study in somewhat different terms, using different criteria and explanations for the identification of multi-word sequences, and so they provide different perspectives on the use of multi-word sequences. For example, some studies describe multi-word sequences that are idiomatic (e.g., expressions like *in a nutshell*), while other studies focus on sequences that are non-idiomatic but perceptually salient (e.g., *you’re never going to believe this*).

A complementary approach, adopted in the present article, is to describe the multi-word sequences that occur most commonly in a given register. We refer to these sequences as ‘lexical bundles’, defined simply as the most frequently recurring sequences of words (e.g., *I don’t know if, I just wanted to*). Lexical bundles are usually not structurally complete and not idiomatic in meaning, but they serve important discourse functions in both spoken and written texts.

Recurrent word sequences have been investigated in several earlier studies, including Salem (1987), Altenberg and Eeg-Olofsson (1990), Altenberg (1998), Butler (1997), Schmitt, Grandage, and Adolphs (2004). The term ‘lexical bundle’ was first used in the Longman Grammar of Spoken and Written English (Biber, Johansson, Leech, Conrad, & Finegan, 1999, chap. 13), which compared the most common recurrent sequences in conversation and academic prose.<sup>1</sup> This framework has been applied in several subsequent studies, including Biber and Conrad (1999), Biber, Conrad, and Cortes (2003, 2004), Cortes (2002, 2004), and Partington and Morley (2004).

<sup>1</sup> The study of lexical bundles reported in the LGSWE (Chap. 12) was based on analysis of the Longman Spoken and Written English Corpus (ca. 4 million words of British English conversation; ca. 3 million words of American English conversation; ca. 5.3 million words of academic prose; see Biber et al., 1999, Chap. 1). The academic prose corpus comprises both academic research articles (ca. 2.7 million words) and advanced academic books (ca. 2.6 million words; see Biber et al., 1999, pp. 32–34). While some advanced academic books can also be used as textbooks, especially in graduate courses, the corpora of academic prose and textbooks were sampled independently: textbooks are mostly written specifically for students, while the articles and books included in the academic prose corpus are written for other professionals.

Biber, Conrad, and Cortes (2004) describe the use of lexical bundles in two university instructional registers: classroom teaching and textbooks. This study also develops a functional framework for the description of lexical bundles in discourse, distinguishing three major functional categories of lexical bundles: stance expressions, discourse organizers, and referential expressions. Stance bundles convey attitudes or assessments of certainty that frame some other proposition. Discourse organizers signal relationships between prior and coming discourse. Finally, referential bundles identify entities or specific parts of entities (see Section 3.2 for a more thorough overview of the functional taxonomy). As a baseline for the description, the 2004 study compared the use of lexical bundles in classroom teaching and textbooks to earlier research on bundles in conversation and academic prose (from Biber et al., 1999). The findings for classroom teaching are especially noteworthy:

- (1) Classroom teaching uses many more bundles than conversation, academic writing, or textbooks.
- (2) Classroom teaching uses many stance bundles, in a similar way to face-to-face conversation, but it also uses many referential bundles, in a similar way to academic writing.
- (3) Classroom teaching actually goes beyond the norms in both directions, using more stance bundles than conversation, and more referential bundles than academic writing.

The (2004) study describes these patterns of use in some detail, interpreting them relative to the complex communicative and situational characteristics of classroom teaching.

The present study extends this line of research by investigating the use of lexical bundles in a wide range of spoken and written university registers: instructional registers (classroom teaching and textbooks), student advising (office hours), ‘management’ registers (class management talk, written course management), institutional registers (service encounters and institutional writing), and student–student academic interactions (study groups).

The study shows that lexical bundles are not an accidental by-product of corpus frequency analysis. Rather, these word sequences turn out to be consistently functional, indicating that high frequency is a reflection of pre-fabricated or formulaic status. Although they go largely unnoticed by speakers, hearers, and analysts, lexical bundles are pervasive in spoken and written texts, where they serve basic discourse functions related to the expression of stance, discourse organization, and referential framing.

However, it is not the case that there is a single pool of lexical bundles that speakers and writers draw from for these discourse functions. Rather, each register employs a distinct set of lexical bundles, associated with the typical communicative purposes of that register. The following sections show that there are fundamentally different sets of lexical bundles associated with spoken university registers in contrast to written registers. Thus, an ESP perspective, considering each register on its own terms, is required to adequately describe the use of lexical bundles in the university context.

The paper is organized as follows: Section 2 introduces the T2K-SWAL Corpus, used as the basis for the present study. Section 3 provides a conceptual introduction to lexical

bundles, together with a description of the methods used to identify them and analyze their primary discourse functions. Section 4 then presents the major findings from the study, comparing the characteristic lexical bundles of each university register, and describing the major discourse functions served by these bundles. Finally, Section 5 discusses conclusions and implications.

## **2. The corpus used for the analysis**

The study is based on a sub-component of the TOEFL 2000 Spoken and Written Academic Language (T2K-SWAL) Corpus. The T2K-SWAL Corpus was designed to represent the range of spoken and written registers that students encounter as part of a university education in the US. The project was sponsored by the Educational Testing Service, with the goal of providing a basis for test construction and validation (see [Biber, Conrad, Reppen, et al., 2004](#)).

The register categories chosen for the T2K-SWAL Corpus are sampled from across the range of spoken and written activities associated with university life, including classroom teaching, office hours, study groups, on-campus service encounters, textbooks, course syllabi, and institutional writing (e.g., university catalog, brochures). Texts from all these registers were sampled from six major disciplines (Business, Engineering, Natural Science, Social Science, Humanities, Education), three levels of education (lower division, upper division, graduate), and four universities (Northern Arizona, Iowa State, California State Sacramento, and Georgia State). The present study includes all major register categories from the corpus, as shown in [Table 1](#).

Classroom teaching, the largest of the spoken registers, includes both lecture-style and more interactive teaching situations. Classroom-management talk occurs at the beginning and end of class sessions, to discuss course requirements, expectations, and past student performance.<sup>2</sup> Office hours are individual meetings between a student and faculty member, for advising purposes or for tutoring/mentoring on academic content. (An office hour ‘text’ usually includes multiple meetings, between one faculty member and the different students who came to her/his office.) Study groups are meetings with two or more students who are discussing course assignments and content. Finally, service encounters are situations where students interact with university staff conducting the business of the university.

The written component of the corpus includes textbooks, course management, and institutional writing. Textbooks include only published books. Written course management includes 10 syllabi ‘text’ files (196 syllabi totaling ca. 34,000 words) and 11 course assignment ‘text’ files (162 individual assignments totaling ca. 18,500 words). Each ‘text’ file combines multiple syllabi or assignments taken from a single academic discipline and level. The main communicative purpose of these texts is similar to classroom management, namely communicating requirements and expectations about a course or particular assignment. As written texts, however, syllabi and assignments differ from classroom management in fundamental ways: they are not interactive, not negotiated, and serve as a kind

---

<sup>2</sup> Classroom-management talk that occurred at other points in a class session (other than the beginning or end) was not extracted from the class session transcripts, and consequently were not included in this sub-corpus.

Table 1

Composition of the corpus for the study (based on the T2K-SWAL Corpus, plus academic prose from the LSWE Corpus)

Register	# of texts	# of words
<i>Spoken</i>		
Classroom teaching	176	1,248,811
Classroom management	40	39,255
Office hours	11	50,400
Study groups	25	141,100
Service encounters	22	97,700
<i>Written</i>		
Textbooks	87	760,619
Course management	21	52,410
Institutional writing	37	151,500
Academic prose <sup>a</sup>	408	5,330,000

<sup>a</sup> The academic prose sub-corpus is taken from the LSWE Corpus.

of formal contract for the course. Finally, institutional writing refers to the miscellaneous campus-related written texts that students encounter. This sub-corpus includes academic program brochures (22,500 words), student handbooks (43,800 words), university magazine articles (2700 words), academic program handbooks (describing program requirements; 27,400 words), and academic policy statements and descriptions taken from university catalogs (52,400 words).

For comparison, we also include the results of an earlier study of lexical bundles in academic prose; the sub-corpus of academic prose texts is taken from the Longman Spoken and Written English Corpus, and the description of lexical bundles in that register is adapted from the Longman Grammar of Spoken and Written English (Biber et al., 1999, Chap. 13).

### 3. Conceptual introduction to lexical bundles

#### 3.1. Operational definition

Lexical bundles are identified using a frequency-driven approach: They are simply the most frequently occurring sequences of words in a sub-corpus of texts from a single register, such as *do you want to* and *I don't know what* in conversation. The frequency cut-off used to identify lexical bundles is somewhat arbitrary. In the present study, we take a conservative approach, setting a relatively high frequency cut-off of 40 times per million words to be included in the analysis. By using a normalized rate of occurrence, we are able to compare the bundles across sub-corpora of different sizes: to be considered a lexical bundle, a four-word sequence must recur at this rate, regardless of the size of the sub-corpus being analyzed. Many of the bundles described here are actually much more common, occurring more than 200 times per million words.

To further limit the scope of the investigation, only four-word sequences are considered here. (The text excerpts in the following sections show that four-word lexical bundles

sometimes occur together to form longer sequences; see Biber et al. (1999, p. 992 ff.) for discussion of these longer lexical bundles.)<sup>3</sup>

A further defining characteristic is that a multi-word sequence must be used in multiple texts to be counted as a lexical bundle; this restriction guards against idiosyncratic uses by individual speakers or authors. Most bundles are distributed widely across the texts in a corpus. For example, even the least common lexical bundles in a large sub-corpus for a single register (such as conversation or classroom teaching) are usually used in at least 20 different texts, while the more common bundles are distributed even more widely.<sup>4</sup>

The specific lexical bundles identified in the present study should be regarded with caution, because the sub-corpora for some registers (e.g., classroom management and office hours) are small for the purposes of lexical bundle analysis.<sup>5</sup> However, the primary goal of the present study is to compare the overall patterns of lexical bundle use across registers: exploring the extent to which bundles are used, and the major functional associations of bundles, in each of these university registers. The T2K-SWAL Corpus and the methods adopted here are suitable for these general purposes.

<sup>3</sup> The quantitative analysis of lexical bundles was undertaken with computer programs that identified and stored every four-word sequence in the corpus. The programs read through each text in the corpus, storing every sequence beginning with the first word of the text and advancing one word at a time. These four-word sequences are candidates for lexical bundle status, but only those sequences that recur at least 40 times per million words, across multiple texts, are regarded as bundles. For example, the first sentence of this footnote consists of the following four-word sequences:

*the quantitative analysis of, quantitative analysis of lexical, analysis of lexical bundles, of lexical bundles was, etc.*

Each time a sequence is identified, it is automatically checked against the previously identified sequences. If a sequence had already been used in the corpus, a running frequency count is computed to show how often the sequence is repeated. In identifying lexical bundles, we relied on orthographic word units, even though these sometimes arbitrarily combine separate words. For example, *into*, *cannot*, *self-control*, and *don't* are all regarded as single words in our analysis. Only uninterrupted sequences of words were treated as lexical bundles. Thus, lexical sequences that spanned a turn boundary or a punctuation mark were excluded. In future research, it would be interesting to extend the research reported here by also considering bundles that span punctuation boundaries.

<sup>4</sup> Additional distributional requirements were used to identify lexical bundles in registers that were represented by small sub-corpora. In those sub-corpora (especially for class management, office hours, and course management), a bundle could occur as few as three times and still have a normed rate of occurrence greater than 40 times per million words. For example, a bundle that occurred only three times in the office hours sub-corpus (with only 50,400 words) would have a normed rate of 60 per million words:

$$(3/50,400) \times 1,000,000 = 60 \text{ per million words}$$

To adjust for this inflated rate of occurrence, an additional restriction was imposed for the analysis of those registers: any bundle with a raw count of 3 must be distributed across three different texts (which also represents three different speakers/writers). Criteria of this type helped to adjust for the differences in representation of the sub-corpora. However, future research should include more comprehensive analyses based on much larger samples, with the sample design more evenly matched across registers.

<sup>5</sup> The frequency of each four-word sequence was normalized to a rate per one million words of text, so the counts for individual lexical bundles are directly comparable across sub-corpora of different sizes. An additional analytical problem arises when comparing the number of lexical bundle types (i.e., different lexical bundles) across sub-corpora. This issue is methodologically complex, because type distributions are not linear (whereas token distributions are); thus it is not possible to directly normalize the number of lexical bundle types to a rate per million words (see, e.g., Biber, 2006; Appendix B).

### 3.2. Structural and functional characteristics of lexical bundles

Lexical bundles have three major characteristics that distinguish them from other kinds of formulaic expressions. First, lexical bundles are by definition extremely common. Second, most lexical bundles are not idiomatic in meaning and not perceptually salient. For example, the meanings of bundles like *do you want to* or *I don't know what* are transparent

---

By using different distributional requirements for the different sub-corpus sizes (see footnote 4), we in effect normalized the distribution of lexical bundle types for the smaller sub-corpora. This can be illustrated by experiments that compare three sub-corpora from classroom teaching:

- (a) a 50,000 word sub-corpus, where we required occurrence in at least three different texts to be considered a lexical bundle;
- (b) a 100,000 word sub-corpus, where we required occurrence in at least four different texts to be considered a lexical bundle;
- (c) a 200,000 word sub-corpus, where we required occurrence in at least five different texts to be considered a lexical bundle.

When lexical bundle analyses were run on these three sub-corpora, we found roughly the same number of lexical bundle types:

- 50,000 word sub-corpus: 37 bundle types,
- 100,000 word sub-corpus: 38 bundle types,
- 200,000 word sub-corpus: 45 bundle types.

For sub-corpora larger than 200,000 words, we found substantially more lexical bundles, because we continued to apply the same requirement that bundles must be distributed in only five different texts. However, the growth rate is much slower than would be predicted from a linear distribution. For example, the full corpus of classroom teaching contains 1.25 million words (6 times larger than the 200,000-word sub-corpus), but it represents only 84 different lexical bundle types (see Fig. 1) – fewer than twice as many as in the 200,000-word sub-corpus.

There are two competing influences here: On the one hand, a larger corpus obviously provides more opportunities for different lexical bundle types. However, at the same time, a lexical sequence must recur more times in a larger corpus to achieve the rate of 40 times per million words. Thus, a sequence that recurs only three times in a 50,000-word corpus has a normalized rate of 60 times per million words. In contrast, a sequence must recur at least 80 times in a 2-million-word corpus to achieve a rate of 40 times per million words. These two competing influences further minimize the differences in the number of lexical bundle types across sub-corpora of different sizes.

Two general patterns emerge from these experiments:

- (1) For the smaller sub-corpora – class management, office hours, study groups, service encounters, course management, and institutional writing – our methods effectively normalize the lexical bundle type counts to a common basis, so these counts are generally comparable across registers. These registers are the primary focus of the present study.
- (2) For the larger sub-corpora – classroom teaching, textbooks, and academic prose – our methods over-estimate the number of lexical bundle types in comparison to the smaller sub-corpora.

This miscalculation – over-estimating the number of bundles in classroom teaching, textbooks, and academic prose – does not influence any of the general conclusions offered in the study. Rather, the differences described in the following sections are actually stronger than they appear. For example, Fig. 1 shows that lexical bundles are by far less common in textbooks and academic prose than in any other university registers. But the results of our experiments indicate that we have over-estimated the number of bundles in these registers in comparison to the number of bundles observed in registers represented by smaller sub-corpora. Thus, the differences among these registers are actually somewhat stronger than shown in Fig. 1.



from the individual words.<sup>6</sup> And finally, lexical bundles usually do not represent a complete structural unit. For example, Biber et al. (1999, pp. 993–1000) found that only 15% of the lexical bundles in conversation can be regarded as complete phrases or clauses, while less than 5% of the lexical bundles in academic prose represent complete structural units. Instead, most lexical bundles bridge two structural units: they begin at a clause or phrase boundary, but the last words of the bundle are the beginning elements of a second structural unit. Most of the bundles in speech bridge two clauses (e.g. *I want to know, well that's what I*), while bundles in writing usually bridge two phrases (e.g., *in the case of, the base of the*).

Although they are neither idiomatic nor structurally complete, lexical bundles are important building blocks in discourse. Lexical bundles provide a kind of pragmatic 'head' for larger phrases and clauses, where they function as discourse frames for the expression of new information. That is, the lexical bundle expresses stance or textual meanings, while the remainder of the phrase/clause expresses new propositional information that has been framed by the lexical bundle. In this way, lexical bundles provide interpretive frames for the developing discourse. For example,

*I want you to write a very brief summary of his lecture.*

*Hermeneutic efforts are provoked by the fact that the interweaving of system integration and social integration [...] keeps societal processes transparent ...*

Three primary discourse functions can be distinguished for lexical bundles: (1) stance expressions, (2) discourse organizers, and (3) referential expressions. Stance bundles express attitudes or assessments of certainty that frame some other proposition. Discourse organizers reflect relationships between prior and coming discourse. Referential bundles make direct reference to physical or abstract entities, or to the textual context itself, either to identify an entity or to single out some particular attribute of the entity as especially important. For each of these primary discourse functions, a taxonomy of sub-categories associated with more specific functions and meanings was developed following a primarily inductive approach. That is, concordance listings of individual bundles were generated to examine the use of each bundle in its discourse context. The bundles were then assigned to groups, and discourse functions for the groups were identified by looking at common traits in the functions of different bundles. Previous theoretical studies of the discourse functions of linguistic features helped inform this final synthesis process. A fuller description of the main discourse functions and of the sub-categories of discourse functions identified within them can be found in Biber, Conrad, and Cortes (2004).

Stance bundles express epistemic evaluations or attitudinal/modality meanings. There are five functional sub-categories of stance bundles: epistemic, desire, obligation, intention/prediction, and ability:

<sup>6</sup> In fact, most longer idioms are far too rare to be considered lexical bundles. Stereotypical idioms such as *kick the bucket* (meaning 'die') and *a slap in the face* (meaning 'an affront') are rarely attested in natural speech or writing. Fiction is one of the few registers that uses idioms and fixed formulas with moderately high frequencies. For example, *kick the bucket* and *a slap in the face* occur around five times per million words in fiction, while they are rarely attested at all in actual face-to-face conversation (see Biber et al., 1999, pp. 1024–1026). Simpson and Mendis (2003) document important pragmatic functions of idioms in classroom teaching, but they found that these expressions are generally rare, and they are often short noun phrases or prepositional phrases (e.g., *the bottom line* and *the big picture*).



**Epistemic lexical bundles:**

*I don't know what the voltage is here*

*There was irony in the fact that the Russian Revolution [...] proclaimed itself to be Marxist ...*

**Desire bundles:**

*I don't want to deliver bad news to her.*

*I want you to take out a piece of paper and jot some notes down ...*

**Obligation (directive) bundles:**

*All you have to do is work on it.*

**Intention/prediction bundles:**

*right now what we're going to take a look at are ones that are [...] positive and beneficial.*

**Ability bundles:**

*I want you to be able to name and define those four curriculum category [sic].*

Discourse organizing bundles function to indicate the overall discourse structure and to signal the informational status of statements. Discourse organizing bundles serve three major sub-functions: topic introduction, topic elaboration/clarification, and referential identification/focus:

**Topic introduction bundles:**

*What I want to do is quickly run through the exercise ...*

**Topic elaboration/clarification bundles:**

*It has to do with the START talks, with the Russians,*

**Identification/focus bundles:**

*For those of you who came late I have the, uh, the quiz.*

(Notice in the above examples that two four-word bundles can occur together, in effect creating a longer five-word or six-word bundle.)

In most cases, identification/focus bundles have a discourse organizing function. These bundles are often used after a lengthy explanation to emphasize or summarize the main point:

*Schizophrenia typically uh will mean that uh separation from reality uh it can mean uh uh you know extreme periods of euphoria and extreme periods of depression it can mean a lot of things – and that's one of the problems of schizophrenia.*

Finally, referential bundles identify an entity or single out some particular attribute of an entity as especially important. Three major sub-categories are distinguished: imprecision indicators, specification of attributes, and time/place/text reference.

**Imprecision bundles:**

*I think really we now have what about, six weeks left in class or something like that.*

**Bundles specifying attributes:**

*It creates a little bit of wealth.*

*These figures give an idea of the size of the ethnological community in Russia.  
 ...students must define and constantly refine the nature of the problem ...*

**Time/place/text-deixis bundles:**

*Children in the United States are not formally employed in farm work ...*

*She's in that.. uh.. office down there.. at the end of the hall ...*

*As shown in Figure 4.4. ...*

#### 4. Lexical bundles in university registers

In the present study, we apply this functional framework to the description of lexical bundles across a range of spoken and written university registers. Fig. 1 displays the distribution of lexical bundle types (i.e., the number of different bundles) in each register. (The overall frequency of lexical bundles in a register is of course much higher, since each bundle type recurs at least 40 times per million words. For example, there are a total of ca. 8000 occurrences of lexical bundles per million words in classroom teaching; see Biber, Conrad, & Cortes, 2004, Fig. 3.)

The patterns of use shown in Fig. 1 are unexpected in several respects. First, within the spoken registers, classroom management and service encounters have the widest variety of bundles. In fact, these two registers use a wider variety of bundles than classroom teaching. This finding is surprising when compared to the Biber, Conrad, and Cortes (2004) study, which concluded that classroom teaching uses an extremely wide variety of bundles (in comparison to conversation, textbooks, and academic prose). In contrast, the present findings suggest that classroom teaching is less distinctive when considered relative to a range of other university registers.

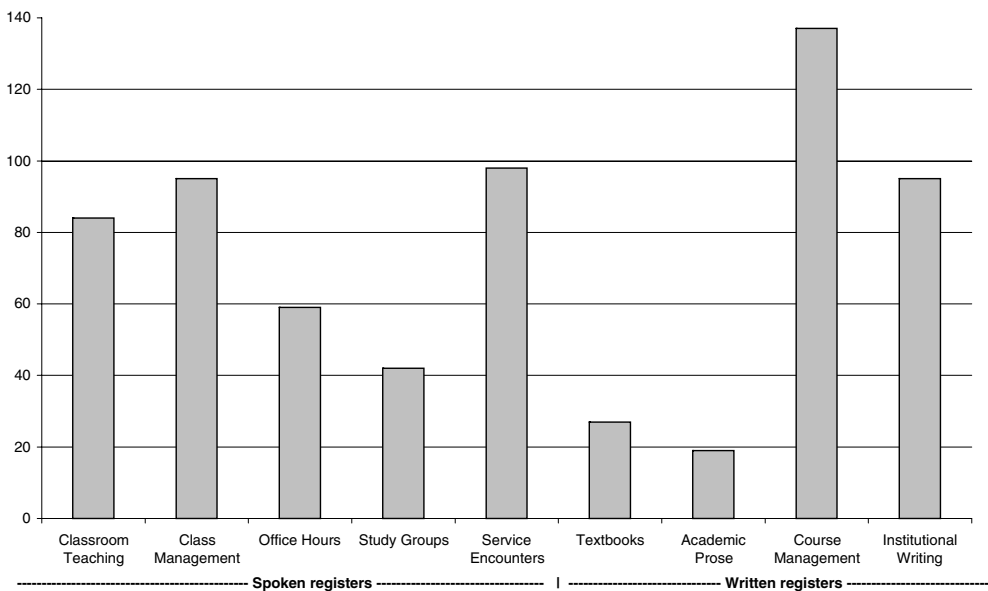


Fig. 1. Number of different lexical bundles (i.e. bundle 'types') across university registers.

Even more noteworthy, though, are the findings for the non-academic written registers. Institutional writing uses as many lexical bundles as any spoken register, while written course management uses more bundle types than any other register considered in the study.

Previous studies have indicated that lexical bundles are generally much more common in spoken discourse than in written discourse (Biber, Conrad, & Cortes, 2004; Biber et al., 1999, chap. 13). The findings in Fig. 1 show that this generalization needs qualification: the extent to which a speaker or writer relies on lexical bundles is strongly influenced by their communicative purposes, in addition to general spoken/written differences. The explanation for the infrequent use of lexical bundles in the academic written registers (textbooks and academic prose) apparently lies in the restricted communicative goals of those registers – focused on informational communication – rather than the written mode per se. We return to a discussion of these overall patterns in Section 5.

#### 4.1. Lexical bundles in spoken university registers

##### 4.1.1. Stance bundles in spoken university registers

Fig. 2 displays the functional distribution of bundles in spoken university registers, showing the number of different lexical bundles (i.e. bundle ‘types’) in each register. For example, there are 84 different lexical bundles in the sub-corpus of classroom teaching,

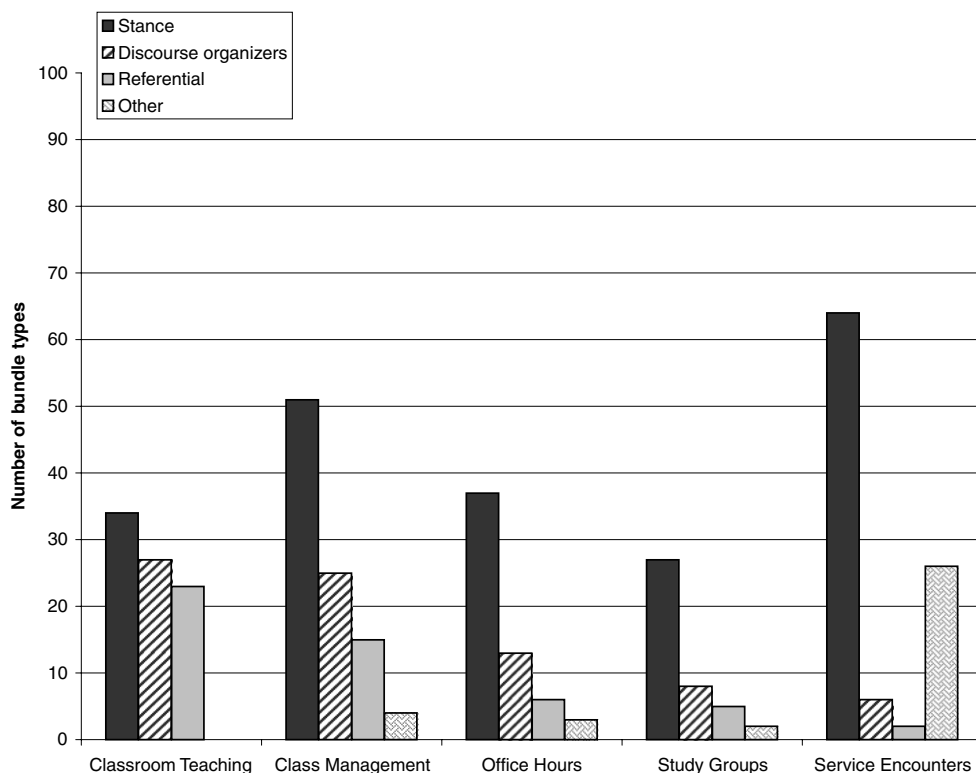


Fig. 2. Functional distribution of lexical bundle types across spoken university registers.

distributed as follows: 34 stance bundles, 27 discourse organization bundles, and 23 referential bundles.

Fig. 2 shows that stance bundles are dominant in the spoken university registers. Stance bundles account for over 60% of the different bundles in office hours, study groups, and service encounters, and they account for well over half of all bundles in class management talk. Only classroom teaching shows a somewhat lesser reliance on stance bundles, but they still account for 40% of the bundle types in this register.

The reliance on stance bundles can be regarded as a general characteristic of spoken university registers. However, Fig. 3 shows that each register relies on different functional sub-categories, reflecting the particular communicative needs of the register. Only one sub-category – obligation/directive bundles – is widely represented in all spoken registers. Epistemic bundles are relatively common in all spoken registers, but they are especially prevalent in study groups. In contrast, the other sub-categories have much more restricted distributions: desire bundles are especially common only in classroom management (and to a lesser extent in service encounters); intention bundles are common only in classroom teaching; and ability bundles are generally rare but moderately common in office hours.

Obligation/directive bundles are the only stance category that is prevalent in all spoken university registers. Even here, though, there are differences in the specific obligation bundles preferred in each register. In classroom management and office hours, we find many

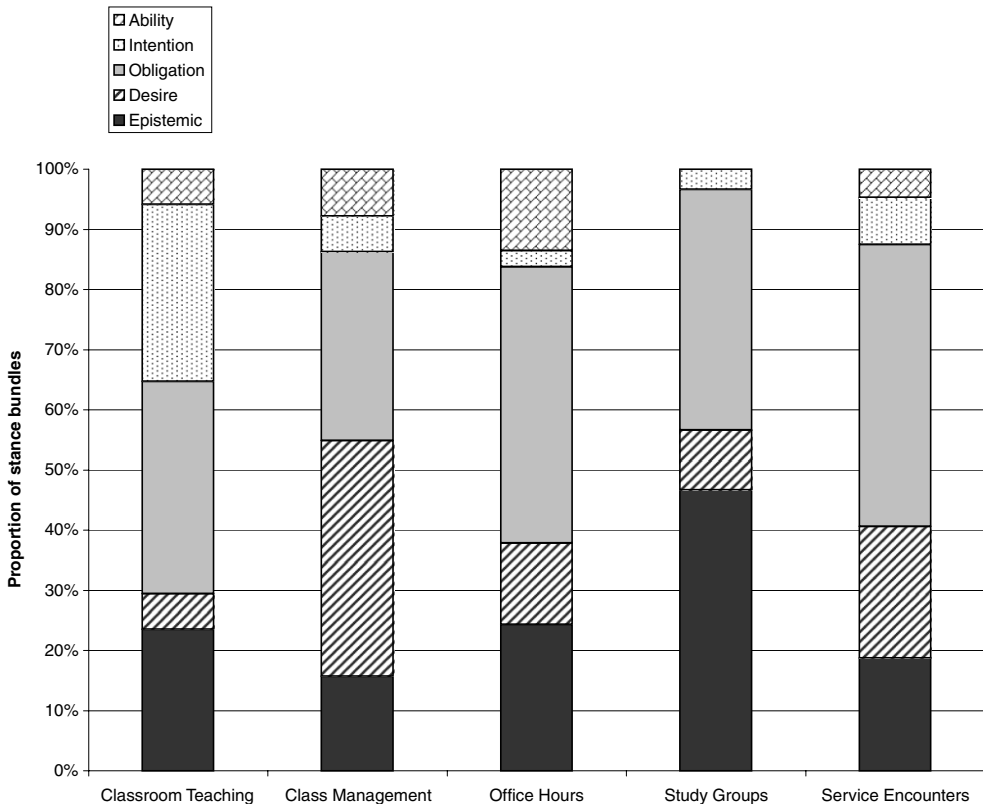


Fig. 3. Proportional breakdown of stance bundles across functional categories: spoken university registers.

obligation bundles with a second person subject. These bundles are found in the discourse of instructors, usually at points where instructors deliver direct, ‘bald-on-record’ instructions about what students must do to complete assignments and other class-related work:

*So if you miss, you just have to find out who was the monitor and go to that person and get whatever you missed.*

*That’s really what you need to do.*

*So you need to go to the library.*

In contrast, obligation bundles in study groups often have the first person plural pronoun ‘we’ as subject: these bundles are used by students to discuss the course requirements that they share:

*we’re gonna have to have time to write this up now.*

*and then we don’t have to worry about this part.*

*I think we need to know those two definitions.*

Obligation stance bundles in university service encounters often have first person ‘I’ as subject, with no mention of ‘you’. These bundles are used by students to identify a required task that they are attempting to complete; for example:

*I need to get some of these copied*

*Do I have to do it in housing?*

In addition, service providers give directives to students that incorporate bundles. These usually include a verb of desire (especially *need*) and refer to expected actions; for example:

*I just need to see a license or something.*

*You need to take it over to enrollment*

Apart from their shared reliance on obligation bundles, spoken university registers differ in the other kinds of stance bundles that they employ. For example, desire bundles are especially common in class management. These are usually used to discuss possible future activities, often taking on an indirect directive function or identifying possible actions that a student could pursue:

*You might want to jot this down for future reference.*

*If you want to talk to them about that you need to talk to the other desk.*

Intention/prediction bundles are common only in classroom teaching. Many of these bundles are used as macro discourse organizers, announcing class topics and activities ahead of time. These bundles are often overtly personal, expressing the speaker’s own intention to perform some future action (often expressed as a joint action). These future actions usually identify the next topic in a class session; for example:

Instructor:

*and in fact what we’re going to do in a minute is move on from this to one that also adjusts for risk as well.*

Other intention/prediction bundles are impersonal, expressing predictions of future events that do not entail the volition of the speaker. These bundles are often used when explaining a logical or mathematical process that involves several steps, as in:

Instructor:

*And so if you require a, twenty percent return on investment, this net present value is going to be zero.*

Office hours use few intention/prediction bundles, but epistemic and ability bundles are more common than in classroom management. Epistemic bundles are often used to hedge claims, as in:

*Instructor: I don't know if it's being offered right now.*

*Instructor: Have you changed the gradient any?*

*Student: I don't think so.*

*Student: so like I just kind of like keep track of them and everything*

Ability bundles are also used in office hours with an indirect directive function, emphasizing the student's ability to accomplish some assigned task, and thereby directing the student to perform the task:

*Instructor: so you should be able to estimate on this when you've got a boundary influence.*

Study groups differ from other spoken university registers in that almost half of all stance bundles have an epistemic function. Similar to office hours, though, these bundles usually function to hedge claims rather than asserting certainty; for example:

*I don't know how to define any of that.*

*I don't know what kind of aggression this one was.*

*I don't think that he terms it socio-emotional.*

*I think it was a review.*

Finally, service encounters have a greater reliance on desire bundles than the other spoken registers. These usually express the desires of the customer:

*I just wanted to pay my fees.*

*Do you want to go back there and check?*

#### 4.1.2. Discourse organizers and referential bundles in classroom teaching

The other two major functional types of bundles are considerably less common than stance bundles in the spoken university registers. Fig. 2 shows that classroom teaching is unique among the university spoken registers in that it relies heavily on discourse organizing bundles and referential bundles, in addition to stance bundles.

Discourse organizing bundles in classroom teaching serve two major functions: topic introduction/focus and topic elaboration/clarification. Topic introduction bundles in classroom teaching provide overt signals to the student that a new topic is being introduced. Many of these incorporate verbs of intention or desire (see Section 4.1.1), but they have the more specialized function of announcing the instructor's intention to begin a new topic:

*But, before I do that, I want to talk about Plato.*

The following example illustrates the use of these bundles for procedural instructions, identifying the major steps in the procedure:

*OK? next thing I want to do is – what I want to do is I want to change the back color [...]* *OK? first thing I want to do is let's set up some colors of the text boxes to start with [...]* *OK? first thing I want to do is let's make the first text box.*

(Notice the use of longer bundles in this excerpt.)

Topic introducing bundles can occur with both first and second person pronouns. The first person plural pronoun *we* and the second person pronoun *you* both invite student participation:

*If we look at um, the California proposition to do away with affirmative action, if you look at the literature, uh this literature basically, is promoting the myth that uh, affirmative action is a racial quota*

The bundle *if you look at* often has a deictic reference, identifying the props required for a topic. This bundle directs student attention to the prop, indirectly introducing a new topic by reference to it:

*And if you look at the spreadsheet, this is what the numbers turn out to be.*

Discourse organizing bundles are also used for topic elaboration or clarification. For example,

*I think there are, are basically two, um, developments that uh, I would hope you to understand how this came about. First has to do with uh, the, changing democratic composition of the central cities ...*

Identification/focus bundles are also relatively common in classroom teaching, focusing on a specific noun phrase following the bundle as especially important. (These bundles were classified as 'referential' rather than 'discourse organizers' in Biber, Conrad, and Cortes (2004); however, because they usually serve a discourse organizing function, they are classified as discourse organizers in the present study). Identification/focus bundles are often used to preview a major topic by stating the main point first:

*I think that one of the things that we're gonna explode is the myth of this harmonious extended family that all lives together in a in a happy go lucky context where everybody loves each other*

In other cases, identification/focus bundles are used after a lengthy explanation to emphasize or summarize the main point:

*And partly that is where, where the disciplines are. You know I think uh, psychology's a much older discipline. It's more developed. They're doing a lot of hypothesis testing, whereas we, you know, we're still working on development of theories so we're looking at how, certain, certain uh, variables are related to others. So we start developing theory. Now, a lot of them haven't been developed but, we're not as advanced as they are in that direction. Anyway, but that's one of the reasons why, the nomenclature changes.*

#### 4.1.3. Longer fixed expressions in service encounters

In addition to regular four-word lexical bundles, there are a number of longer fixed expressions that occur with high frequencies in service encounters. These are composed of four-word bundles that occur in an overlapping sequence, as in:



*it's going to take just a minute for that to go through go ahead and sign that (for me (please))*

These longer fixed expressions were found only in service encounters, reflecting the routines that are characteristic of this speech event. Providers in service encounters repeat the same actions several times each day, and they use formulaic interactions accompanying those actions: greeting customers, asking if they can help, completing transactions, etc. For example, the following interaction occurs repeatedly in the student business services office. There are minor variations, depending on the particular circumstances, but the overall structure of the interaction is relatively fixed:

*Provider: hello. hi.*

*Student: I need to pay this*

*Provider: OK. [types on keyboard] seven oh eight ... OK it's going to take just a minute for that to go through*

*Provider: alright*

*[...]*

*Provider: OK go ahead and sign that for me. there you go*

*Student: thanks*

*Provider: have a good day*

*Student: you too*

These fixed expressions in service encounters are structurally complete and thus differ from typical lexical bundles. In fact, these longer fixed expressions are often used as an entire turn in a service encounter. It is noteworthy that such structurally complete formulaic expressions are found with high frequencies in only one university register: service encounters, which is constrained by a highly restricted set of actions and topics.

#### 4.2. Lexical bundles in written university registers

Fig. 4 presents the functional distribution of lexical bundle types in the written university registers. Lexical bundles are generally rare in textbooks and academic prose (see Biber, Conrad, & Cortes, 2004). In contrast, both course management and institutional writing use a large number of lexical bundle types. However, the functional distribution of bundles is strikingly different in these two registers.

In written course management, over half of all bundle types are stance bundles. Referential bundles are also relatively common in this register. There are fewer discourse organizers in written course management, although even this category is common in comparison with the other written registers.

In institutional writing, almost 70% of all bundle types are referential; stance bundles and discourse organizers are considerably less common in this register. (Referential bundles are also dominant in academic writing (textbooks and academic prose), even though bundles are relatively rare overall in those registers.)

Place references are the most common type of referential bundle in institutional writing. This is not surprising given the need to refer repeatedly to offices and other institutions on campus (e.g., *in the college of, from the office of*). Many of these word sequences are names or titles of an institution, rather than the typical lexical bundles found in other registers; for example:

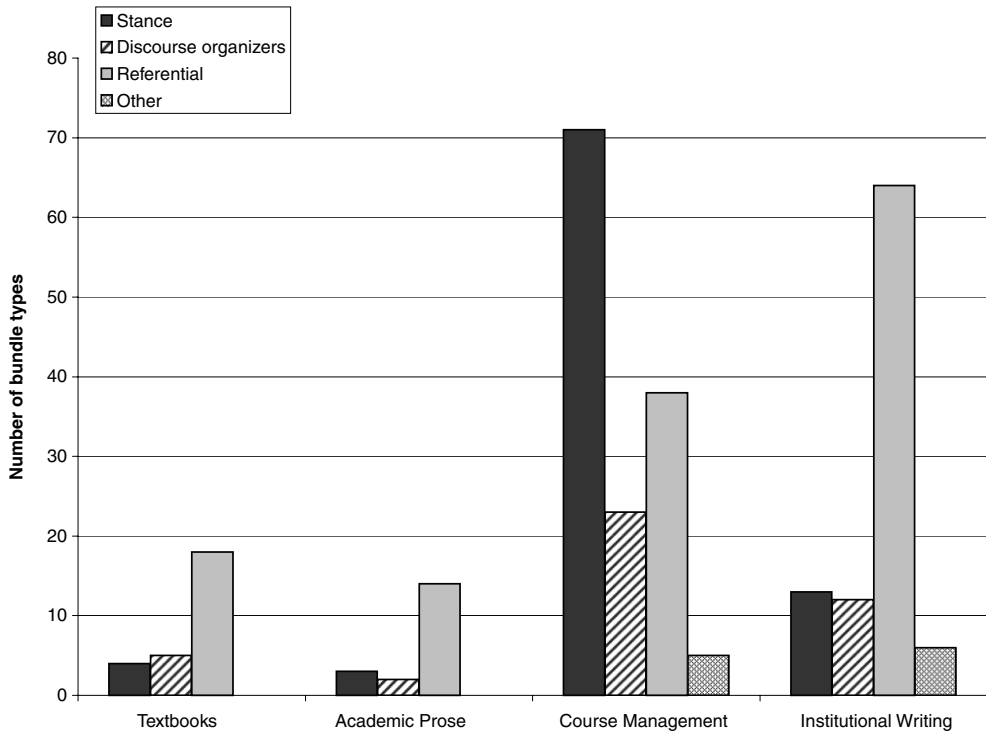


Fig. 4. Functional distribution of lexical bundle types across written university registers.

*at Georgia State University, Immigration and Naturalization Service, the office of residence life*

Time bundles are also relatively common in both course management and institutional writing. Many of these refer to specific times that are especially relevant to university life, such as class periods or semesters:

Course management:

*beginning of each class, during my office hours, over the course of, the beginning of each, the end of each*

Institutional writing:

*at the time of, fall and spring semesters, the first day of, the semester in which*

Intangible framing attributes are more prevalent in syllabi and other written course management materials, being used to introduce course content and the conceptual organization of a course:

*Architecture 271 is an introduction to the concept of fit in architecture.*

*An additional aim of this course is to investigate the nature of the ethnographic enterprise itself.*

*We shall analyze the relationship between the data they gather and the claims they make.*

Discourse organizing bundles are also relatively common in course management, where they indicate the overall organization and goals of a course:

*Instead, our goal is to UNCOVER the meaning of key themes in psychology and grow in understanding and applying these themes to our world.*

*The objective of this course is to expose the student to introductory material relating to electromagnetics.*

#### 4.2.1. Stance bundles in written course management

Spoken classroom management and written course management are similar in that they both use many different stance lexical bundles (compare Figs. 2 and 4). However, as Fig. 5 shows, these two registers tend to rely on different functional sub-categories: Both registers use obligation bundles, but spoken classroom management also relies heavily on desire bundles (see discussion above), while written course management employs intention/prediction bundles. A closer consideration of these bundles in context reveals that most of them tend to be used to express directives, regardless of their subcategory.

Although obligation bundles are common in both spoken and written management registers, the specific bundles used in the two registers have different structural characteristics. In spoken class management, obligation bundles usually contain an active voice verb that expresses modal necessity, such as *you + have to/need to/want to* (see Section 4.1.1). In contrast, obligation bundles in written course management usually contain a passive voice verb (e.g., *students are expected to, you are encouraged to*). These bundles function to explicitly state the requirements or expectations of a course:

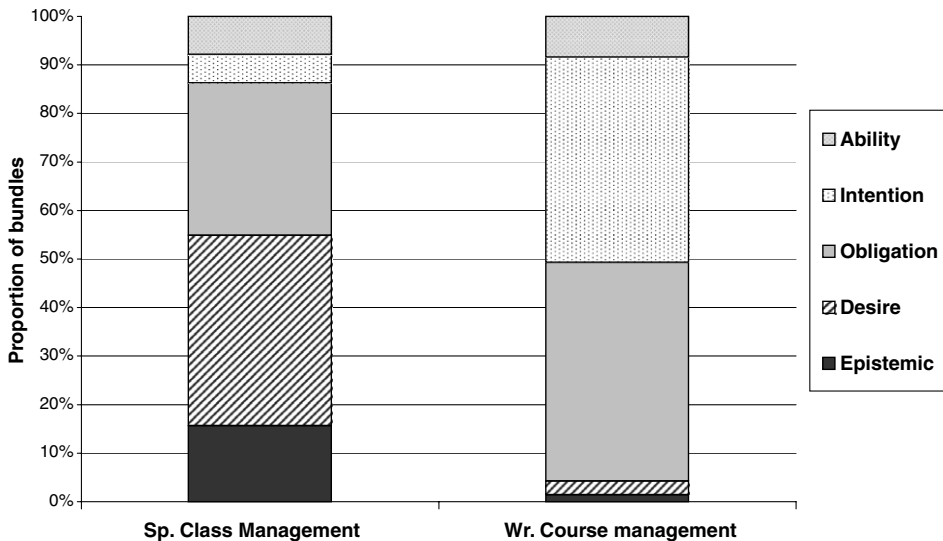


Fig. 5. Proportional breakdown of stance bundles across functional categories: spoken classroom management vs. written course management (syllabi, etc.).

*You are expected to attend the classes.*  
*Each student will be required to present a 40 minute talk.*

Intention/prediction bundles are also common in written course management. Some of these introduce the topics or organization of a course:

*We will look at relevant data and file structures and incorporate database access through embedded SQL.*

However, intention/prediction bundles more often have directive functions in classroom management, simply stating future requirements and expectations, such as:

*All students will be required to defend their work in oral examinations.*  
*Homework will be assigned weekly.*

In sum, spoken classroom management and written course management use different sets of lexical bundles, with different structural characteristics, but for similar communicative purposes. In both these registers, lexical bundles are used primarily for explicit directives expressing obligation. Both registers also commonly use lexical bundles for indirect directives, but they prefer different functional types: spoken class management favors desire bundles, while written course management prefers intention/prediction bundles. In speech, the instructor expresses directives by telling students what he/she ‘wants’ or ‘would like’; in writing, the instructor often expresses directives by identifying events that will occur in the future, usually with a passive voice verb and no indication of the agent.

It is interesting that written course management goes beyond the norms of all spoken university registers (including classroom management) in its dense use of lexical bundles. This pattern in part reflects the relatively constrained purposes of course management writing, particularly course syllabi. In the United States, the syllabus fulfils important legal and administrative functions: it is used as a contract between the instructor and the students, and as a permanent record for accountability and documentation purposes (e.g., for accreditation of academic programs and institutions, to determine course equivalency in transferring situations, etc.) (Eberly, Newton, & Wiggins, 2001; Parkes & Harris, 2002). Because of their status as contracts and permanent records, syllabi typically include the same types of information or policies, such as course schedule, grading policies, attendance policies, academic dishonesty policy, etc. Thus syllabi have become highly conventionalized, addressing a restricted set of communicative concerns, and as a result they have evolved to rely heavily on a large set of lexical bundles expressing those different concerns.

#### 4.3. *Summary and discussion*

Several major distributional patterns emerge from the preceding distributional analyses of lexical bundles in university registers, including:

- Lexical bundles are used to some extent in all university registers.
- Lexical bundles are common in all spoken university registers, regardless of purpose.
- Lexical bundles are relatively rare in the academic/instructional written registers.

- At the other extreme, lexical bundles are more common in the written non-academic registers than in any other university register.

The functional analyses revealed several additional general patterns:

- In the spoken registers, lexical bundles are most commonly used to express stance.
- Stance is also the most common function of bundles in student ‘management’ registers, spoken and written.
- Discourse organizing functions are relatively important in the more ‘teacher-centered’ spoken registers.
- In general, referential functions are the dominant use of lexical bundles in the written university registers.
- Each of these registers uses a particular set of lexical bundles representing specific functional sub-classes. For example:
  - Intention stance bundles are especially common in classroom teaching and in written course management.
  - Desire stance bundles are especially common in class management.
  - Epistemic stance bundles are especially common in study groups.
- Written course management is exceptional: it makes frequent use of bundles from all three functional classes (although stance bundles are especially common).

The findings show the importance of both mode (bundles are common in all spoken registers) and communicative purpose (bundles are common in all student management registers). In contrast, previous research on university registers has found that mode differences are by far the most important in accounting for linguistic variation: spoken university registers are consistently different from written university registers in the use of a wide range of lexical and grammatical features (see, e.g., Biber, 2006, chap. 8). For example, verbal and clausal features are common in all spoken university registers and relatively rare in all written registers. In contrast, complex noun phrase features are common in all written university registers and relatively rare in all spoken registers. The results of previous ‘multi-dimensional’ analyses similarly show a fundamental divide between the spoken and written university registers (see, e.g., Biber, Conrad, Reppen, Byrd, & Helt, 2002): all spoken university registers, regardless of purpose, are ‘involved’, ‘situated in reference’, and characterized by the absence of ‘impersonal’ styles; in contrast, all written university registers, again regardless of purpose, are highly ‘informational’, ‘elaborated in reference’, and marked for ‘impersonal’ styles.

The present study suggests that lexical bundles are fundamentally different from other lexico-grammatical features in their patterns of use: grammatical features are influenced primarily by physical mode (speech versus writing), while the use of lexical bundles is influenced by both mode and communicative purpose.

The student management registers most clearly illustrate the difference in use between grammatical features and lexical bundles. Biber (2006) shows that spoken class management and written course management are strikingly different with respect to most lexico-grammatical features. For example, class management is similar to other spoken registers in relying heavily on verbal, clausal, and pronominal features; written course management is similar to textbooks and institutional writing in relying heavily on complex noun phrase constructions. In contrast, the study here has shown that spoken and written

student management registers are similar in their dense use of lexical bundles. At the same time, these two registers differ in their specific sub-classes of bundles, reflecting their origins in speech versus writing.

In summary, the study has identified three basic patterns: the pervasiveness of lexical bundles in university language; the strong functional basis of lexical bundles; and the extent to which each register is associated with a distinctive set of bundles, serving particular discourse functions.

## 5. Conclusion

Lexical bundles are not the only kind of fixed or formulaic language; idioms and formulaic expressions (e.g., *happy birthday*; *keep off the grass*) have actually been much more widely studied. In fact, researchers like Wray (2002, chap. 2) discount the value of studying formulaic language using frequency criteria. On the one hand, many formulaic sequences are not frequent; stereotypical idioms, for example, are generally rare (see Biber et al., 1999, pp. 1024–1026). But at the same time, Wray (2002) claims that not all frequent word sequences are formulaic; that a sequence could be frequent but not pre-fabricated (see, e.g., p. 31).

To our knowledge, this last assertion has not been demonstrated or even empirically investigated. In fact, there have been few studies that directly investigate whether any word sequences are truly *pre-fabricated*: ‘stored and retrieved whole from memory ...’ (Wray, 2002, p. 9). Because we lack adequate experimental methods to study the structure and content of the mental lexicon, researchers have instead relied on the linguistic properties of word sequences, identifying pre-fabricated sequences based on intuition or perceived salience, rather than on evidence from actual linguistic production and comprehension. In practice, pre-fabricated sequences are usually restricted to those word combinations that are relatively fixed in form and not compositional in meaning – i.e. idiomatic to some extent (see, e.g., Howarth, 1998; Hudson, 1998). Extremely frequent word sequences do not meet these criteria, and thus they have been typically disregarded in earlier research.

However, detailed studies of the discourse functions of lexical bundles – defined on a frequency basis – indicate that these word combinations can also be regarded as pre-fabricated units. Lexical bundles do not express new propositional meaning; rather, they are used to provide a discourse frame for some other new assertion. They are not idiomatic in meaning, but they do serve important discourse functions related to the expression of stance, discourse organization, and referential framing.

Thus, in contrast to earlier theoretical frameworks, we hypothesize that high frequency is a reflection of pre-fabricated status. While these word sequences are generally not idiomatic, the fact that they are consistently functional suggests that they are stored whole in memory and used as unanalyzed language chunks for textual or interpersonal discourse functions. Of course, this is an empirical research question; we need future experimental research to determine if native speakers actually comprehend and produce lexical bundles as unanalyzed pre-fabricated sequences (see also Schmitt et al., 2004).<sup>7</sup>

---

<sup>7</sup> When testing the psychological status of lexical bundles, it is important to match the register of the experimental setting with the characteristic register of the target bundles. For example, conversational bundles should not be tested in formal written contexts, and academic writing bundles should not be tested in spoken contexts.

Regardless of their psycholinguistic status, lexical bundles are clearly useful devices for the comprehension and construction of discourse. There is no denying the fact that these sequences are pervasive in university discourse, and we hope that we have also demonstrated that they are consistently functional. A student in the university context will frequently encounter these word sequences; failure to understand their textual and interpersonal functions will obviously influence her/his success in dealing with both spoken and written language situations. We do not yet know whether students have difficulty understanding the intended functions of these sequences, but it seems uncontroversial that learners of English must control these expressions to succeed in university contexts.

Although researchers generally recognize the importance of formulaic language for native-like competence and fluency (see Van Lancker-Sidtis & Rallon, 2004; Wray & Perkins, 2000, among many), there is little agreement as to how formulaic sequences might be effectively introduced in L2 teaching curricula. For example, Wray (2000) reviews three major pedagogical frameworks (Lewis, 1993; Nattinger & DeCarrico, 1992; Willis, 1990) but finds them all inadequate to some extent.

Because they are so frequent, we might assume that students will naturally acquire an understanding of the discourse roles of lexical bundles, and they would therefore not need to be overtly taught. Research in SLA has argued for the importance of both frequency and salience as two basic parameters influencing the acquisition of language features (see, e.g., Ellis, 2002; Goldschneider & DeKeyser, 2001). However, some researchers have argued that factors such as perceptual salience and developmental readiness are more important than frequency for acquisition (see Gass & Mackey, 2002, for a review). For example, there is evidence to suggest that 'marked' forms (which are generally salient but not necessarily frequent) are sometimes acquired before 'unmarked' alternatives (e.g., Bardovi-Harlig, 1987). If these findings are generalizable, they might indicate that the discourse functions of lexical bundles are not easily acquired: they are not perceptually salient, despite their high frequency. If so, lexical bundles might be good candidates for overt instruction.

However, this is also a research question that must be investigated empirically in future classroom-based research. We need studies that investigate the extent to which students come to control the discourse functions of lexical bundles through natural acquisition, as well as studies that test the effectiveness of pedagogical approaches for the overt instruction of lexical bundles. Such studies should investigate whether lexical bundles can be successfully taught, and if so, what gains in comprehension and fluency of production result from the instruction. At present, we do not have answers to these questions. But what we do have is clear linguistic evidence that high-frequency lexical sequences are not accidental or mere artifacts of computational analysis; rather, these sequences serve important discourse functions in both spoken and written registers. As such, they provide a necessary complementary perspective to descriptions of formulaic language based solely on perceptual criteria.

## References

- Altenberg, B. (1998). On the phraseology of spoken English: the evidence of recurrent word-combinations. In A. Cowie (Ed.), *Phraseology: Theory, analysis and applications* (pp. 101–122). Oxford: Oxford University Press.
- Altenberg, B., & Eeg-Olofsson, M. (1990). Phraseology in spoken English. In J. Aarts & W. Meijs (Eds.), *Theory and practice in corpus linguistics* (pp. 1–26). Amsterdam: Rosopi.
- Atkinson, D. (1996). The philosophical transactions of the Royal Society of London, 1675–1975: a sociohistorical discourse analysis. *Language in Society*, 25, 333–371.



- Bardovi-Harlig, K. (1987). Markedness and salience in second-language acquisition. *Language Learning*, 37, 385–407.
- Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. Amsterdam: John Benjamins.
- Biber, D., & Conrad, S. (1999). Lexical bundles in conversation and academic prose. In H. Hasselgard & S. Oksefjell (Eds.), *Out of corpora: Studies in honor of Stig Johansson* (pp. 181–189). Amsterdam: Rodopi.
- Biber, D., Conrad, S., & Cortes, V. (2003). Lexical bundles in speech and writing: an initial taxonomy. In A. Wilson, P. Rayson, & T. McEnery (Eds.), *Corpus linguistics by the lute: A festschrift for Geoffrey Leech* (pp. 71–92). Frankfurt/Main: Peter Lang.
- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at . . . : lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25, 371–405.
- Biber, D., Conrad, S., Reppen, R., Byrd, P., & Helt, M. (2002). Speaking and writing in the university: a multi-dimensional comparison. *TESOL Quarterly*, 36, 9–48.
- Biber, D., Conrad, S., Reppen, R., Byrd, P., Helt, M., Clark, V., et al. (2004). *Representing language use in the university: Analysis of the TOEFL 2000 spoken and written academic language corpus*. ETS TOEFL monograph series, MS-25. Princeton, NJ: Educational Testing Service.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *The Longman grammar of spoken and written English*. London: Longman.
- Butler, C. S. (1997). Repeated word combinations in spoken and written text: some implications for functional grammar. In C. S. Butler, J. H. Connolly, R. A. Gatward, & R. M. Vismans (Eds.), *A fund of ideas: Recent developments in functional grammar* (pp. 60–77). Amsterdam: IFOTT, University of Amsterdam.
- Cortes, V. (2002). Lexical bundles in academic writing in history and biology. Doctoral dissertation, Northern Arizona University.
- Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: examples from history and biology. *English for Specific Purposes*, 23, 397–423.
- DeCarrico, J., & Nattinger, J. R. (1988). Lexical phrases for the comprehension of academic lectures. *English for Specific Purposes*, 7, 91–102.
- Eberly, M. B., Newton, S. E., & Wiggins, R. A. (2001). The syllabus as a tool for student-centered learning. *The Journal of General Education*, 50(1), 56–74.
- Ellis, N. (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition*, 19, 91–126.
- Ellis, N. C. (2002). Frequency effects in language processing. A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24, 143–188.
- Flowerdew, J. (Ed.). (2002). *Academic discourse*. New York: Longman.
- Gass, S., & Mackey, A. (2002). Frequency effects and second language acquisition. *Studies in Second Language Acquisition*, 24, 249–260.
- Goldschneider, J. M., & DeKeyser, R. M. (2001). Explaining the “natural order of L2 morpheme acquisition” in English: a meta-analysis of multiple determinants. *Language Learning*, 51, 1–50.
- Hewings, M. (Ed.). (2001). *Academic writing in context: Implications and applications*. Birmingham: The University of Birmingham Press.
- Howarth, P. (1996). *Phraseology in English academic writing*. Tübingen: Max Niemeyer Verlag.
- Howarth, P. (1998). Phraseology and second language proficiency. *Applied Linguistics*, 19, 24–44.
- Hudson, J. (1998). *Perspectives on fixedness: Applied and theoretical*. Lund: Lund University Press.
- Hyland, K. (1998). *Hedging in scientific research articles*. Philadelphia: John Benjamins.
- Khuwailah, A. A. (1999). The role of chunks, phrases and body language in understanding co-ordinated academic lectures. *System*, 27, 249–260.
- Lewis, M. (1993). *The lexical approach*. Hove: Teacher Training Publications.
- Lindemann, S., & Mauranen, A. (2001). “It’s just real messy”: the occurrence and function of just in a corpus of academic speech. *English for Specific Purposes*, 20, 459–475.
- Markkanen, R., & Schroder, H. (Eds.). (1997). *Hedging and discourse: Approaches to the analysis of a pragmatic phenomenon in academic texts*. Berlin: Walter de Gruyter & Co.
- Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical phrases and language teaching*. Oxford: Oxford University Press.
- Parkes, J., & Harris, M. B. (2002). The purposes of a syllabus. *College Teaching*, 50(2), 55–61.
- Partington, A., & Morley, J. (2004). From frequency to ideology: investigating word and cluster/bundle frequency in political debate. In B. Lewandowska-Tomaszczyk (Ed.), *Practical applications in language and computers – PALC 2003* (pp. 179–192). Frankfurt a. Main: Peter Lang.

- Poos, D., & Simpson, R. (2002). Cross-disciplinary comparisons of hedging: some findings from the Michigan Corpus of Academic Spoken English. In R. Reppen, S. Fitzmaurice, & D. Biber (Eds.), *Using corpora to explore linguistic variation* (pp. 3–21). Philadelphia: John Benjamins.
- Powell, C., & Simpson, R. (2001). Collaboration between corpus linguists and digital librarians for the MICASE web interface. In R. Simpson & J. Swales (Eds.), *Corpus linguistics in North America* (pp. 32–47). Ann Arbor: University of Michigan.
- Salem, A. (1987). *Pratique des segments répétés*. Paris: Institut National de la Langue Française.
- Schmitt, N. (Ed.). (2004). *Formulaic sequences*. Amsterdam: John Benjamins.
- Schmitt, N., Grandage, S., & Adolphs, S. (2004). Are corpus-derived recurrent clusters psycholinguistically valid? In N. Schmitt (Ed.), *Formulaic sequences* (pp. 127–152). Amsterdam: John Benjamins.
- Simpson, R. (2004). Stylistic features of spoken academic discourse: the role of formulaic expressions. In U. Connor & T. Upton (Eds.), *Discourse in the professions: Perspectives from corpus linguistics* (pp. 37–64). Amsterdam: John Benjamins.
- Simpson, R., Briggs, S. L., Ovens, J., & Swales, J. M. (2002). *The Michigan corpus of academic spoken English*. Ann Arbor, Michigan: The Regents of the University of Michigan.
- Simpson, R., & Mendis, D. (2003). A corpus-based study of idioms in academic speech. *TESOL Quarterly*, 37, 419–441.
- Swales, J. M. (2004). *Research genres: Exploration and applications*. Cambridge: Cambridge University Press.
- Van Lancker-Sidtis, D., & Rallon, G. (2004). Tracking the incidence of formulaic expressions in everyday speech: methods for classification and verification. *Language and Communication*, 24, 207–240.
- Weinert, R. (1995). The role of formulaic language in second language acquisition: a review. *Applied Linguistics*, 16, 180–205.
- Willis, D. (1990). *The lexical syllabus*. London: Harper Collins.
- Wray, A. (2000). Formulaic sequences in second language teaching: principle and practice. *Applied Linguistics*, 21(4), 463–489.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.
- Wray, A., & Perkins, M. (2000). The functions of formulaic language: an integrated model. *Language and Communication*, 20, 1–28.