# *On the other hand*: Lexical bundles in academic writing and in the teaching of EAP

**PAT BYRD**

*Georgia State University*

**AVERIL COXHEAD**

*Victoria University of Wellington*

## ABSTRACT

Corpus linguistics has demonstrated that language-in-use involves repetition of fixed and semi-fixed multiword combinations. Language-in-use also involves the use of formulaic patterns that can run from one word to many words. Currently, much of the reported research focuses on *lexical bundles.* To find out how lexical bundles function across different disciplinary areas in universities, the analysis reported here started with the creation of a list of lexical bundles used in arts, commerce, law, and science (each made up of seven subject areas) in a corpus written academic English. The use of the bundles in each of these four disciplinary areas was analyzed and compared to published results of similar data. This process led to a short but powerful list of 21 four-word lexical bundles that occur across these four disciplines. The discussion of the results of this search for widely used lexical bundles leads to a consideration of challenges in taking lexical bundle data into the English for Academic Purposes (EAP) classroom. The

Address for correspondence: Averil Coxhead, School of Linguistics & Applied Language Studies, Victoria University of Wellington, Wellington 6140, New Zealand; Email: averil.coxhead@vuw.ac.nz

*University of Sydney Papers in TESOL*, 5, 31-64.

©2010 ISSN: 1834-3198 (Print) & 1834-4712 (Online)

challenges lead to suggestions (a) for teachers about working with word lists made up of multiword sequences and (b) for researchers about providing data that would be especially useful for EAP teachers and students.

## REPEATED LEXICAL SEQUENCES

Corpus linguistics has demonstrated that language in use is characterized by repetition of fixed and semi-fixed multiword combinations and by use of formulaic patterns. These patterns can run from one word to many words. They include, at least, frames such as *the … of the …*, idioms, collocational pairs, and sets of two or more contiguous words. The ability to recognize and to produce such patterns is thought to be of importance for language learners to develop both fluency and appropriate usage for particular settings. Thus, many studies are now reporting on high frequency multiword sets (e.g., Baker, 2006; Biber & Barbieri, 2007; Biber, Conrad, & Cortes, 2004; Cortes, 2004; Hyland, 2008). Currently, much of the reported research focuses on *lexical bundles.*

*Lexical bundles* combine three or more words that are repeated without change for a set number of times in a particular corpus. The definition of lexical bundle also requires that the bundle must occur widely in the texts that make up the corpus. This requirement avoids sets that are just characteristic of a particular speaker or writer rather than broadly used by a discourse community. The lexical bundle is discovered by having a software program find all of the set phrases of a certain length in a certain range of texts in the corpus. The program then reports back on the frequency of the sets that are found. Cut off points are decided by the researcher based on what seems reasonable given the volume of data. A tradition is developing to base decisions on what is reported in similar studies. Thus, lexical bundle studies report high levels of frequency rather than statistical significance. Biber, Johansson, Leech, Conrad, and Finegan (1999) used a cut off of 10 occurrences per million words. Biber (2006), Cortes (2004), and Hyland (2008) require that lexical bundles occur at least 20 times per million words. Biber and Barbieri (2007) raise the limit to 40 occurrences per million words for the phrases to be analyzed in that study of spoken and written university language.

Lexical bundles have been used to analyze characteristic language for a variety of communicative types and purposes. Biber *et al.* (1999) differentiated among newspaper prose, academic writing, conversational English, and fiction. Cortes (2004) investigated differences between student and published writing in history and biology. Hyland (2008) analyzed the lexical bundles in a corpus made up of samples of published writing along with student writing in dissertations and master's theses to investigate differences among disciplines.

Previous research demonstrates that particular discourse types are characterized by the use of grammar and vocabulary somewhat differently from the use of language in other discourses (e.g., Baker, 2006; Biber *et al.*, 2004; Hyland, 2008; Pickering & Byrd, 2008; Stubbs & Barth, 2003). A research report in biology will be instantly recognizable as different from a chapter in an introductory undergraduate textbook in biology. Even an introductory chapter in a biology textbook will share grammar and some vocabulary and perhaps even a stylistic preference with a biology research report. Such an introductory textbook sample, however, is likely to share features with other introductory textbooks. This similarity is because of their shared purposes in introducing entry-level undergraduate students to basic content and terminology aimed at a particular age group of new university students. Teasing out the patterns of language that are found across different university disciplines and those that are restricted to use in particular disciplinary areas remains an important task for applied corpus linguistics.

There is little advice based on solid research on the most useful pedagogical approach to lexical bundles and phrases (Coxhead, 2008a). Granger and Meunier (2008, p.249) state there is an "urgent need for more empirical evidence of the actual impact of a phraseological approach to teaching and learning." A number of other challenges face teachers and learners of EAP when it comes to lexical bundles. These challenges include the limited numbers of bundles in single texts and deciding how words lists of bundles might be used to guide principles and decisions for teaching and learning. Another problem is deciding what to do when shorter bundles occur within longer ones, for example *at the end of* contains both *at the end* and *the end of*. Teachers and students need information on the use of bundles in context. They also require convincing arguments as to why it is worth spending time on potentially well-known words such as *result* in the bundle *as a result of*. This is especially true considering that students may not encounter these bundles often in their reading and listening. We consider these challenges of using lexical bundles in EAP courses after the discussion and make suggestions on how teachers might work with lexical bundles in their classrooms.

## CORPUS AND METHODS

The study reported here uses the corpus created for the development of the Academic Word List (Coxhead, 2000). The AWL was developed using a written academic corpus of 3.6 million running words with four academic disciplines: arts, commerce, law, and science. Each of these disciplines contained seven subject areas. The corpus contained 414 texts. Table 1 gives the subject areas for each of the disciplinary sub-areas of the total corpus. In the selection of texts for the corpus, an attempt was made to balance the number of short texts (2,000 - 5,000 running words), medium length texts (5,000 - 10,000 running words) and long texts (over 10,000 running words) between the four faculty areas as much as possible.

These four areas were chosen because they represent four major areas of study for undergraduate students at Victoria University of Wellington, New Zealand, as well as at other similar educational systems outside of New Zealand. Each of seven subject areas contained approximately 875,000 running words. For example, science contained biology, chemistry, computer science, geography, geology, mathematics, and physics (see Table 1). Each of the seven subject areas contains approximately 125,000 running words. Table 2 gives the word counts for each of the sub-disciplines.

**TABLE 1**
**Subject areas in the disciplinary sub-section of the academic corpus**

| Arts | Commerce | Law | Science |
|---|---|---|---|
| Education | Accounting | Constitutional | Biology |
| History | Economics | Criminal | Chemistry |
| Linguistics | Finance | Family and Medico-Legal | Computer Science |
| Philosophy | Industrial Relations | International | Geography |
| Politics | Management | Pure Commercial | Geology |
| Psychology | Marketing | Quasi-Commercial | Mathematics |
| Sociology | Public Policy | Rights and Remedies | Physics |

**TABLE 2**
**Percentage of total words in lexical bundles by disciplinary sub-corpus**

| Discipline | Total wrds subcorpus | Four-word LBs 20/mil types | 4-wrd LBs 20/mil tokens | # wrds LBs 20/mil | % total words |
|---|---|---|---|---|---|
| Arts | 909663 | 106 | 3265 | 13060 | 1.44 |
| Commerce | 921496 | 198 | 6094 | 24376 | 2.65 |
| Law | 898209 | 376 | 12206 | 48824 | 5.44 |
| Science | 927634 | 112 | 3396 | 13584 | 1.46 |

The corpus contained journal articles, book chapters, course workbooks, laboratory manuals, and course notes. Some sections of the Lancaster-Oslo/Bergen (LOB) (Johannson, 1978), Brown (Kucera & Francis, 1967), Wellington (Bauer, 1993) corpora were also included, drawing criticism from Hyland and Tse (2007) because of the age of the LOB and Brown corpora in particular. The sections from these corpora comprise approximately 6% percent of the total written academic corpus. Since the development of the corpus, one duplicate text has been found in the chemistry section. This text contains 16,608 running words out of a total of 131, 494, reducing the

Chemistry section used in this paper to 114,886 running words and the total number of texts in the corpus used for the current study to 413. Word counts in this current study were developed using Wordsmith Tools 5.0 (Scott, 2006, see www.lexically.net/wordsmith /). The original AWL study (Coxhead, 2000) used the Range program (Heatley, Nation & Coxhead, 2002). Because the two programs use different approaches to identifying words, the counts here are somewhat different from those in the original study.

Our analysis started with the creation of a list of lexical bundles used in all four disciplines of the AWL corpus. Then, the frequency of each lexical bundle in each of the four areas was analyzed, seeking patterns of similarity and difference across the disciplines. Because of small variations in the sizes of the four areas (see Table 2), the frequency data for the lexical bundles were standardized per million words. We selected bundles that appeared at least 20 times per million words in the standardized data. After a discussion of the results of this search for widely used lexical bundles, we consider challenges in taking lexical bundle data into the EAP classroom. We also present a series of suggestions for teachers working with word lists made up of multiword sequences and for researchers providing data that would be especially useful for EAP teachers and students.

## RESULTS AND DISCUSSION

Table 2 gives the percentage of total words in lexical bundles for each disciplinary area in the AWL corpus. Arts and Sciences have the fewest words in lexical bundles. Law is at the other extreme with 5.44% of the total words in set phrases. That is, they are repeated without variation in the wording. Commerce is in the middle with 2.65% of the total words in four-word set phrases. Seventy-three of the bundles are shared across all four disciplines (each bundle occurring at least 20 times per million words). The 73 bundles are used 9,904 times for a total of 39,616 words out of the 3.6 million words in the AWL corpus. That is, this seemingly small set of highly frequent and widely used bundles makes up 1.1% of the total words in the AWL corpus.

However, the 73 lexical bundles that are shared by all four areas do not occur in equal numbers in each of the disciplines. An analysis reported in Table 3 reduces the list to those bundles reasonably well distributed across the four disciplines by selecting only those bundles that make up at least 10% of each discipline. For example, *on the other hand* occurs 353 times in the combined AWL corpus; 23% of those are in Arts, 27% in Commerce, 35% in Law; and 15% in Science. This process reduced the number of shared bundles to 35 that are highly likely to be found in all four disciplinary areas in the AWL corpus (see Table 3).

**TABLE 3**
**Percentages of lexical bundles widely used in four disciplinary areas of the AWL corpus**

| **Shared lexical bundles** | **Freq** | **Arts raw** | **Arts %** | **Com raw** | **Com %** | **Law raw** | **Law %** | **Sci raw** | **Sci %** |
|---|---|---|---|---|---|---|---|---|---|
| On the basis of | 380 | 115 | 30 | 87 | 23 | 133 | 35 | 45 | 12 |
| On the other hand | 353 | 81 | 23 | 97 | 27 | 122 | 35 | 53 | 15 |
| As a result of | 283 | 65 | 23 | 64 | 23 | 109 | 39 | 45 | 16 |
| The end of the | 281 | 74 | 26 | 58 | 21 | 93 | 33 | 56 | 20 |
| At the end of | 235 | 48 | 20 | 62 | 26 | 66 | 28 | 59 | 25 |
| At the same time | 230 | 73 | 32 | 69 | 30 | 54 | 23 | 34 | 15 |
| The nature of the | 229 | 34 | 15 | 51 | 22 | 103 | 45 | 41 | 18 |
| In the form of | 217 | 36 | 17 | 94 | 43 | 59 | 27 | 28 | 13 |
| In terms of the | 213 | 64 | 30 | 56 | 26 | 41 | 19 | 52 | 24 |
| In the absence of | 211 | 37 | 18 | 48 | 23 | 94 | 45 | 32 | 15 |
| At the time of | 185 | 28 | 15 | 44 | 24 | 90 | 49 | 23 | 12 |
| As well as the | 150 | 48 | 32 | 46 | 31 | 23 | 15 | 33 | 22 |

**Note:** Freq = Frequency Com = Commerce Sci = Science

## TABLE 3
## Percentages of lexical bundles widely used in four disciplinary areas of the AWL corpus (continued)

| **Shared lexical bundles** | **Freq** | **Arts raw** | **Arts %** | **Com raw** | **Com %** | **Law raw** | **Law %** | **Sci raw** | **Sci %** |
|---|---|---|---|---|---|---|---|---|---|
| It is clear that | 121 | 30 | 25 | 25 | 21 | 39 | 32 | 27 | 22 |
| In the United States | 119 | 22 | 18 | 32 | 27 | 49 | 41 | 16 | 13 |
| That there is a | 119 | 47 | 39 | 21 | 18 | 30 | 25 | 21 | 18 |
| The way in which | 119 | 51 | 43 | 25 | 31 | 30 | 37 | 13 | 16 |
| Is likely to be | 114 | 33 | 29 | 42 | 37 | 25 | 22 | 14 | 12 |
| It is possible to | 113 | 17 | 15 | 28 | 25 | 28 | 25 | 40 | 35 |
| It is important to | 108 | 29 | 27 | 23 | 21 | 21 | 19 | 35 | 32 |
| As part of the | 104 | 31 | 30 | 23 | 22 | 35 | 34 | 15 | 14 |
| In the same way | 101 | 15 | 15 | 16 | 16 | 40 | 40 | 30 | 30 |
| That there is no | 100 | 25 | 25 | 20 | 20 | 42 | 42 | 13 | 13 |
| It is difficult to | 96 | 18 | 19 | 27 | 28 | 35 | 36 | 16 | 17 |
| The case of the | 96 | 24 | 25 | 30 | 31 | 27 | 28 | 15 | 16 |
| It is necessary to | 93 | 18 | 19 | 23 | 25 | 19 | 20 | 33 | 35 |
| A result of the | 90 | 21 | 23 | 25 | 28 | 35 | 39 | 9 | 10 |
| A wide range of | 87 | 37 | 43 | 16 | 18 | 14 | 16 | 20 | 23 |
| The relationship between the | 87 | 15 | 17 | 37 | 43 | 25 | 29 | 10 | 11 |
| The rest of the | 86 | 21 | 24 | 19 | 22 | 21 | 24 | 25 | 29 |

**Note:** Freq = Frequency Com = Commerce Sci = Science

**TABLE 3**
**Percentages of lexical bundles widely used in four disciplinary areas of the AWL corpus (continued)**

| **Shared lexical bundles** | **Freq** | **Arts raw** | **Arts %** | **Com raw** | **Com %** | **Law raw** | **Law %** | **Sci raw** | **Sci %** |
|---|---|---|---|---|---|---|---|---|---|
| The development of the | 83 | 25 | 30 | 24 | 29 | 20 | 24 | 14 | 17 |
| Is one of the | 82 | 21 | 26 | 23 | 28 | 22 | 27 | 16 | 20 |
| In addition to the | 81 | 18 | 22 | 23 | 28 | 20 | 25 | 20 | 25 |
| From time to time | 79 | 10 | 13 | 34 | 43 | 27 | 34 | 8 | 10 |
| In a number of | 75 | 16 | 21 | 15 | 20 | 31 | 41 | 13 | 17 |
| In the presence of | 75 | 14 | 19 | 16 | 21 | 9 | 12 | 36 | 48 |

**Note:** Freq = Frequency Com = Commerce Sci = Science

All together these 35 bundles are used a total of 5195 times in the AWL corpus. At four words per bundle, these 35 add up to 20,780 words or approximately 0.58% of the 3.6 million words in the whole corpus. These figures demonstrate the utility of this relatively small set of words. Let's now compare these results with other lists of highly frequent lexical bundles in academic prose.

Several researchers, including Biber, Conrad, and Cortes (2004) and Hyland (2008), have published lists of frequent lexical bundles. We compared those lists with the list of the 35 shared and highly frequent bundles in the AWL corpus. We found a list of bundles (see Table 4) that can be viewed by teachers and materials writers as highly important and fairly stable across a variety of types of academic prose.

## TABLE 4
## Shared lexical bundles found in AWL corpus; Biber, Conrad & Cortes (2004), and Hyland (2008)

| **Lexical bundles** | **Freq in AWL** | **All 3** | **AWL & Biber *et al.*** | **AWL & Hyland** |
|---|---|---|---|---|
| On the basis of | 380 | *** | | |
| On the other hand | 353 | *** | | |
| As a result of | 283 | *** | | |
| The end of the | 281 | *** | | |
| At the end of | 235 | *** | | |
| At the same time | 230 | *** | | |
| The nature of the | 229 | *** | | |
| In the form of | 217 | *** | | |
| In terms of the | 213 | *** | | |
| In the absence of | 211 | | ** | |
| At the time of | 185 | | ** | |
| As well as the | 150 | *** | | |
| In the United States | 119 | | ** | |
| The way in which | 119 | | ** | |
| It is possible to | 113 | | ** | |
| It is important to | 108 | | ** | |
| It is necessary to | 93 | | ** | |
| The relationship between the | 87 | | | ** |
| The rest of the | 86 | | ** | |
| Is one of the | 82 | *** | | |
| In the presence of | 75 | | ** | |

## ANALYZING THE STRUCTURAL FEATURES OF ACADEMIC PROSE

The 21 lexical bundles in our final list were analyzed (Table 5) using the structural categories in Biber (2006). As has been shown in other studies of academic prose (Biber, 1988, 2006), the structural features of these shared lexical bundles indicate the importance of long, complex noun phrases in such writing.

**TABLE 5**
**Grammatical structures of 21 widely used lexical bundles**

| **Structure** | **Count** | % |
|---|---|---|
| prepositional phrase + of | 8 | 38 |
| other prepositional phrase | 4 | 19 |
| noun phrase + of | 3 | 14 |
| anticipatory it | 3 | 14 |
| noun phrase + other complement | 2 | 10 |
| be + complement | 1 | 5 |
| Total | 21 | 100 |

For example, the subject of the following sentence from the AWL corpus runs for 19 words: *A director who has inside information material to the assessment of the value of the company's shares or securities may transact those shares*. Academic prose is considered to be 'noun-centric'. Our analysis supports this finding because we found only a limited use of verbs in the 21 highly frequent lexical bundles. The only lexical verbs found on the list are forms of *be*: for example, *that there is a, it is clear that,* and *is likely to be*. Additionally, particular prepositional phrases are highly frequent adverbials: *in the case of, on the basis of, on the other hand, as a result of*, and others.

A lack of passive bundles among these shared phrases coincides with Hyland (2008) and other studies which have suggested that passive voice might be more characteristic of science than of other academic disciplinary areas. This is a finding that supports Conrad

(2008) in her observation that passive voice is not necessarily a highly frequent form in all academic prose (see also Tarone, Dwyer, Gillette & Icke, 1981). A shared list that combines lexical bundles from arts, commerce, and law along with science is likely to reduce the importance of a structural feature which is more characteristic of one of the disciplines than the others.

The bundles have a heavy use of complex prepositional phrases (*as well as the*) and of prepositional phrases as post noun modifiers (*the value of the*). This finding suggests that the study of English prepositions at intermediate and more advanced levels should move beyond work with concrete adverbial meanings that are traditional in ESL/EFL (e.g., *in Atlanta)*. Such lessons could focus on the use of adverbials such as *in the case of* or *on the basis of* which are likely to appear in any of the disciplinary areas in which students might study.

## Discourse functions of the shared lexical bundles

Various systems have been developed to analyze the discourse functions of particular types of prefabricated language in context (see e.g., Biber, Conrad, & Cortes, 2003; Sinclair & Mauranen, 2006). Whatever terms are used, these systems generally include three basic categories: "presentation of content" and "organization of the discourse/text," and "expression of attitudes by the writer/speaker." Within each of these sub-sets, a variety of related discourse functions can be clustered. While the system proposed in Biber, Conrad, and Cortes (2003) is being widely adopted, Hyland (2008) demonstrates the usefulness of creating analytical systems closely adapted to the nature of the discourse from which the bundles are taken. Systems developed for research purposes can seem overly complex. They might also use terminology that is not easily understood by less-proficient language learners when they are taken into classroom settings as part of the curriculum or teaching materials.

## Lexical bundles for presenting and discussing content

We have limited our initial analysis of these highly frequent lexical bundles to a rough division of the bundles into their most essential purposes in (a) the presentation of content, (b) the creation of connections within the text, and (c) the expression of attitudes by the writer. Limiting our analysis is an attempt to provide a system that teachers might find more directly applicable to teaching EAP. Table 6 shows the results of that analysis.

**TABLE 6**
**Functional analysis of shared lexical bundles in AWL corpus, Biber *et al.* (2004) and Hyland (2008)**

| **Lexical bundles** | **Freq in AWL** | **Presenting & discussing content** | **Organizing discourse** | **Expressing attitudes** |
|---|---|---|---|---|
| On the basis of | 380 | √ | | |
| On the other hand | 353 | | √ | |
| As a result of | 283 | √ | | |
| The end of the | 281 | √ | √ | |
| At the end of the | 235 | √ | | |
| The nature of the | 229 | √ | | |
| At the same time | 230 | √ | √ | |
| In terms of the | 213 | √ | | |
| In the form of | 217 | √ | | |
| In the absence of | 211 | √ | | |
| At the time of | 185 | √ | | |
| As well as the | 150 | √ | | |
| In the United States | 119 | √ | | |
| The way in which | 119 | √ | | |
| It is possible to | 113 | | | √ |
| It is important to | 108 | | | √ |

**TABLE 6**
**Functional analysis of shared lexical bundles in AWL corpus, Biber et al. (2004) and Hyland (2008) (continued)**

| Lexical bundles | Freq in AWL | Presenting & discussing content | Organizing discourse | Expressing attitudes |
|---|---|---|---|---|
| It is necessary to | 93 | | | √ |
| The relationship between the | 87 | √ | | |
| The rest of the | 86 | √ | | |
| Is one of the | 82 | √ | | |
| In the presence of the | 75 | √ | | |

These categories are not absolute and bundles can have overlapping functions. For example, the concordance lines for the bundles reveal that 17 of the 281 uses of *the end of the* are textual organizers. They point the reader to a section of a paper or book chapter. The rest of the uses primarily indicate an end time for some event or process. A few of the uses are formulaic in nature: *end of the road, end of the matter, at the end of the day*, but most uses are completed by the addition of a time to indicate the whole period being discussed.

Two bundles have considerable overlap in use: about half (113/235) of the uses of *at the end of* are in the context of the five-word bundle *at the end of the.* Both bundles are primarily used for presenting and discussing content.

Most of the uses of *at the same time* indicate two events or process that occur simultaneously. However, the relationship is less about time than about simultaneity, as in this example: *They are told they must participate in development but at the same time not forget their true kodrat* (*nature, destiny, duty*). The phrase is also used to indicate a logical relationship between events or process. It comes very close to being a discourse organizer with a meaning similar to *however*, as in this example: *…the adjustment process has not run its full course. At the same time, a number of early results are reasonably clear.*

The bundle *as well as the* works at the sentence level to connect two related concepts, groups, or events in sentences, for example: … *requires rigorous testing by the vendor as well as the accepting organization*. The implication is that the unit that comes after the bundle is the basic, standard one. That is, testing is naturally done by the accepting organization but should also be done by the vendor. Thus, the phrase is conceptually complex for the reader/writer. For a teacher, this type of lexical bundle raises other issues: bundles are often incomplete units that the user completes for particular uses, adding the basis to *on the basis of*.

Knowing what to do with the remnant of a noun phrase such as *the* in *as well as the* seems more of a challenge for teachers than completing a prepositional phrase. One implication is that the more powerful lexical bundle might be *as well as* with *as well as the* as a related sub-type. In the AWL corpus, *as well as* occurs 709 times and is used in all four disciplinary areas. It is one of the most frequent combinations in the AWL corpus totaling about .02% of the words. Thus, the four-word bundle might better be taught within the context of the three-word bundle and as an extension of the bundle.

**Lexical bundles to express writer attitudes**

The short list of lexical bundles used to express the writer's attitude suggests two things that might be useful for EAP teachers. First, English has a much larger set of these attitude markers (see Hunston & Thompson, 2000; Biber, Conrad & Cortes, 2004), so the list could be expanded to include others that might be found in lists such as those provided by Biber *et al.* (1999). Second, writers of academic prose use a variety of tools, including hedges (Hyland, 1998), so that they can indicate an emotional stance toward the content of their writing. For second language writers, learning appropriate ways to use these stance markers will pose cultural as well as linguistic challenges. They might be met by careful reading of appropriate academic texts to learn how other writers make use of such bundles. For readers, it is important to recognize these stance markers as part of the critical reading process.

### Lexical bundles to organize discourse

Among the discourse organizers on the list of bundles shared across the disciplines, *on the other hand* is particularly interesting because it frequently appears independent from its traditional partner, *on the one hand.* In only 39 of the 100 uses of *on the one hand* is the word *other* found nearby. Thus, *on the other hand* is most often used as a transition and contrast marker without the prior use of *on the one hand*. While this study focuses on written Academic English, a similar pattern is suggested by frequency of these bundles in MICASE, the corpus of spoken academic English provided by the University of Michigan. In MICASE, *on the one hand* is used 33 times while *on the other hand* is used 66 times (Pickering & Byrd, 2008). This overlap in usage could enhance the opportunities for learning since students could work with authentic samples of both spoken and written academic English using the same high frequency lexical bundle.

## POSSIBLE LIMITS ON THE USEFULNESS OF THE LEXICAL BUNDLE

When we realize that we are finding exactly the same words repeated in exactly the same order over and over again in many different texts by many different writers from many different disciplinary backgrounds, lexical bundles demand our attention. On the one hand, what seems like a surprisingly high frequency of a small set of these words suggests their importance for language learners. Yet, on the other hand, high frequency lexical bundles do not make up a dominant percentage of the corpora so far reported in published research. Hyland (2008) found that lexical bundles used at least 20 times per million words made up 2% of the words in his corpus, a percentage that is slightly larger than our finding of 1.1% of the AWL corpus.

At the same time, the scale used to report lexical bundles is typically in terms of the number of bundles per million words. For example, *on the basis of* (Table 3) occurs 308 times in the 3.6 million words that make up the AWL corpus. That's 106 times per million words, or 53 times per 500,000 words, or twice per 15,625 words. Studies of vocabulary acquisition report that learners need many encounters with a word or phrase before it becomes part of their lexicon (Nation, 2008). Few learners will read a million words in an EAP class. Most will read fewer than the 15,000 words needed to encounter *on the basis of* even twice.

Additionally, lexical bundles are just one of a variety different types of prefabricated and often repeated language. Counts of the percentage of preformulated/formulaic language in English run as high as 25-80%. Altenberg (1998) is often quoted for his estimate that some 80% of the words in the London-Lund Corpus of Spoken English "form part of a recurrent word-combination in one way or another." Other similar estimates of the high percentage of language in preformulated phrases include, for example, Erman (2007), Erman and Warren (2000), Pawley and Syder (1983), Sinclair and Mauranen (2006), and Wray (2008).

If such estimates of the percentage of texts made up of often-repeated set phrases are close to the mark, then the question that arises is "if a written academic corpus contains 25% or more of its words in prefabricated or formulaic language and if high frequency lexical bundles make up only 1-2% of that language, what kinds of units make up the rest?" The answer seems to be that any authentic sample of English is going to include a wide range of formulaic language. In addition to lexical bundles, a text would include at least frequent repetition of two-word collocations, frames with slots, some commonly used metaphoric language, and possibly an idiom or two. It will also include sets of technical vocabulary that are characteristic of a particular field of study.

## LIMITS OF STUDYING A PARTICULAR TEXT

Lexical bundles and other formulaic linguistic patterns are features of language that are revealed by study of large corpora. The whole range of formulaic language which is characteristic of a language cannot be discovered by studying a small single-authored sample of the language. Similarly, no single text will involve all elements of repeated language characteristic of the larger corpus. However, teachers and students work with single texts as part of the educational process. They often study from single journal articles and single textbook chapters chunk by chunk through the time allowed for a particular course. Indeed, many EAP courses use only pieces of whole publications. For example, they might use part of a journal article or part of a textbook chapter. As we consider the status of lexical bundles within the context of teaching EAP, we wonder "What are teachers and students likely to find when they study a particular text?"

To explore the use of lexical bundles at the text level, we analyzed a chapter from a widely used textbook on biology-ecology (Campbell & Reece, 2005). This chapter contained nearly 14,000 words. To check the general lexical features of the chapter, we ran it through Tom Cobb's Vocabulary Profiler on the Compleat Lexical Tutor website (n.d.) (see Table 7). While we are not interested in single words in this study, the profile below shows that the chapter has the general features of vocabulary found in most academic writing (Biber, 1988; Coxhead, 2000, 2008b).

TABLE 7
Vocabulary profiler data on Cambell and Reece (2005)

| | Families | Types | Tokens | Percent |
|---|---|---|---|---|
| K1 Words (1-1000): | 593 | 1001 | 8775 | 63.48% |
| Function: | ... | ... | (4860) | (35.16%) |
| Content: | ... | ... | (3915) | (28.32%) |
| > Anglo-Sax<br>=Not Greco-Lat/Fr Cog: | ... | ... | (1495) | (10.81%) |
| K2 Words (1001-2000): | 185 | 282 | 623 | 4.51% |
| > Anglo-Sax: | ... | ... | (196) | (1.42%) |
| 1k+2k | | ... | ... | (67.99%) |
| AWL Words (academic): | 284 | 460 | 1241 | 8.98% |
| > Anglo-Sax: | ... | ... | (58) | (0.42%) |
| Off-List Words: | ? | 728 | 1937 | 14.01% |
| | 1062+? | 2470 | 13824 | 100% |

We generated a list of four-word phrases that were used at least two times in the chapter using Wordsmith Tools 5.0 (Scott, n.d.). We compared that list to the shared lexical bundles reported in Table 4. Then we compared the chapter-specific lexical bundles to those in the AWL science sub-corpus and to Hyland's list (2008) for his biology data.

We found that of the 35 lexical bundles that were shared across the four disciplines in the AWL corpus, two (*in the United States* and *it is important to*) were used in the chapter by (Campbell & Reece, 2005). Of the four-word lexical bundles found for the science discipline of the AWL corpus, four lexical bundles were also used in Campbell and Reece (2005). They are, *for example in the, is an example of, the total number of,* and *which of the following*. Hyland (2008) lists 50 four-word lexical bundles that he found in his biology samples. This list includes phrases shared across his corpus and some that were found only in the biology sub-corpora. None of the bundles restricted to Hyland's biology data were used in the biology-ecology textbook chapter analyzed here.

Perhaps an EAP teacher might turn to working with bundles that are frequent in texts that students will study. This way, a teacher might avoid bundles from corpus studies which are likely not to be frequent in a particular text. To evaluate that approach, we used Wordsmith Tools 5.0 to find 143 four-word clusters in the chapter by Campbell and Reece (2005). Most of these were used just two times in the chapter. Bundles that were used more frequently involved discipline-specific terminology: *the declining population approach* (8 repetitions) and *the red cockaded woodpecker* (6 repetitions). Eight other bundles appeared 4 times in the chapter; 16 appeared 3 times; many of these are content specific such as *a biodiversity hot spot* and *biology and restoration ecology*. However, these bundles are possibly specific to this text or to specialization in this particular sub-area of biology. If the purpose of the EAP course is to prepare students for that specialized work, then these lexical bundles could be useful. And if study of a variety of texts from that specialization confirms the importance of these bundles, then study of these sets might reward student effort. But words found in a single text might not necessarily be a guide for anything other than that particular text. On the other hand, if the purpose is to prepare students for a broader encounter with academic English, then these specialized bundles might be useful for reading the text but not necessarily useful for having students study for use in other contexts. This point leads us to our next section on the challenges EAP teachers face when considering how they might use lexical bundles in their classrooms.

## CHALLENGES IN USING LISTS OF LEXICAL BUNDLES IN EAP COURSES

We have found six challenges for teachers in using lexical bundle data in EAP. We look at each of these challenges below. Where possible, we draw on examples and analysis from our study of these lexical bundles.

## Challenge 1: Working with word lists of bundles published in research reports

Word lists have elicited a mixture of responses from teachers (see Folse, 2004). However, as Tom Cobb (n.d.) says, "Learners like word lists, so let's give them good ones." The same can be said of lists of bundles. Lists of bundles can be used as the basis for materials design and curriculum development, as Jones and Haywood (2004) outlined in their study of teaching lexical bundles in an EAP course. Textbooks and dictionaries are beginning to show new levels of awareness of the importance of bundles, phraseology and collocations, but in the case of textbooks, more so at intermediate than advanced levels (Gouverneur, 2008). When adopting or adapting lists of bundles, teachers and learners need to know how a list has been developed. For example, was the list derived from written and spoken corpora? What kinds of texts were included in the corpus? Are they representative of the reading of undergraduate or postgraduate learners? What principles of selection were used? How has the list been evaluated?

Perhaps a both/and approach would be useful at this stage, whereby lists are used to guide the selection of lexical bundles and texts are used to provide context and support for instruction and decision making. Cortes (2004) found that lexical bundles were not used very often in writing by history and biology students. Rather these students tend to rely on a small number of bundles which are often not used in the same way as professionals within the fields use them. Teachers and learners would benefit from the findings of such studies.

## Challenge 2: The length of lexical bundle to teach when shorter bundles are reported inside longer ones

Researchers often decide to report bundles of a particular length with four-word bundles being especially popular. Generally, the core reason for the decision is that longer bundles are not as frequent as shorter ones and thus provide for more manageable analysis of large corpora. Additionally, a small subset of shorter bundles can be assumed to be captured as part of the longer ones. The discussion of *as well as the* given above illustrates the problem for teachers of this reasonable research decision. That is, what is the better teaching/learning unit: *as well as* or *as well as the*? The list of highly frequent shared bundles (see Table 4) includes seven of these shorter bundles folded into longer ones: *the end of the, the nature of the, in terms of the, as well as the, the relationship between the, the rest of the,* and *is one of the*. A comparison (see Table 8) shows that these 3-word bundles are folded into the four-word forms 21% to 64% of the total use of the 3-word bundles. For example, *the end of* is used 499 times with *the end of the* making up 281 of those occurrences.

**TABLE 8**
**Relationship between four-word and 3-word lexical bundles**

| **Shared Lexical Bundles** | **AWL Total** | **four-word as a % of 3-word use** |
|---|---|---|
| the end of the | 281 | 56% |
| the end of | 499 | |
| the nature of the | 229 | 50% |
| the nature of | 460 | |
| in terms of the | 213 | 26% |
| in terms of | 816 | |
| as well as the | 150 | 21% |
| as well as | 709 | |
| the relationship between the | 87 | 36% |
| the relationship between | 242 | |
| the rest of the | 86 | 64% |
| the rest of | 134 | |
| is one of the | 75 | 51% |
| is one of | 146 | |

When a four-word bundle is an especially frequent combination even when compared to the 3-word version, a teacher might choose to focus on the four-word version. In other instances, the 3-word version seems to need to take priority. As with other reported corpus data, the problem for teachers is getting access to such data about related lexical bundles.

### Challenge 3: Lack of information on use in context of bundles in published lists

Teachers and students need more detailed information about the use of these highly frequent lexical bundles in the context of academic prose. For example, we analyzed the concordance lines in which *on the basis of*, the most frequent shared lexical bundle, (see Table 7) is used. We found three patterns of use for this bundle.

a. ***Used at the beginning of a sentence:*** In this use, *on the basis of* functions both to provide a transition and to specify methods or data used to carry out a process. This use needs an extended context to show how the phrase transitions and justifies as shown in this example from the AWL corpus:

> *... Clyne's research provides valuable information on the distribution of a large number of these languages in Australia (Clyne, 1985, 1991, Clyne and Kipp, 1996).* ***On the basis of*** *his analyses, Clyne also identifies a number of "unequivocally important" factors as relevant in accounting for different rates of language shift in different communities....*

***b. Used as an adverbial of reason in a passive sentence or clause to explain the way that a decision was made or data handled:***

> *Meanwhile, unskilled and unassisted migrants, most notably from Southern Europe, were accepted* ***on the basis of*** *nomination by relatives in Australia....*

Verbs that were used at least 3 times in this pattern include *calculated* (3), *claimed* (3), *classified* (3), *decided* (3), *developed* (5), *justified* (3), *made* (9 uses), *selected* (4), and *targeted* (4). These can be divided into two sets based on meaning (Table 9), those focused on how something was developed or those focused on how something was used or applied.

**TABLE 9**
***On the basis of* for information development or use**

| Data or information development | Data or information use |
|---|---|
| ascertained | accepted |
| calculated | cannot be assumed |
| classified | challenged |
| conducted | claimed |
| determined | considered |
| developed | criticized |
| quantified | decided |
| | demonstrated |
| | expected |
| | judged |
| | justified |
| | learnt |
| | made |
| | selected |
| | targeted |

Other verbs that are used only 1 or 2 times provide variations on these uses…in other words, while these individual verbs are not used often, the meaning pattern is an important one.

c. ***Meaning strengthened or diminished with an adverbial:*** *apparently, largely, normally, only, partly, primarily, purely, simply,* and *solely.*

*Only for L. notosaurus was the decision on its specific distinction made* ***solely on the basis of*** *allopatric data****.***

However, teachers will find it difficult to get access to information about the contexts in which the lexical bundles are used because much published research involves analysis of privately held corpora.

### Challenge 4: Lack of face validity for some EAP students

One of the difficulties is the face validity of teaching bundles, for example, *as a result of,* to EAP learners who may already be undertaking undergraduate or postgraduate study at university. This kind of focus on well known words such as *result* may seem remedial at best or a waste of time at worst. Often the difficulties around bundles (like *as a result of*) are to be found in the dense academic language to the left and the right of the bundle itself. They could also be found in the lack or overuse of such a bundle in learners' own academic writing. Also, the examples in the section above indicate that there may be other much more difficult features of academic prose (such as *allopatric data*) than the lexical bundle *as a result of.*

### Challenge 5: Contradiction between analytical approach in teaching and use as unanalyzed chunks

Wray (2002) argues there is a basic contradiction between teaching formulaic sequences by pulling them apart in language classrooms and the way that speakers use the sequences in an unanalysed way. Putting lexical bundles back together post analysis and using them accurately and appropriately in speaking and writing are not easy tasks for learners. This is particularly true when they are participating in short courses of instruction.

### Challenge 6: Having students read enough text to encounter the lexical bundles frequently enough for learning

Finally, as we have seen above in the figures reporting on the number of occurrences of lexical bundles in academic texts, a large amount of reading is required to encounter bundles in context.

Students who are preparing for academic study need to read academic texts rather than focusing extensively on stories or other literary types (Coxhead, 2006). It is important to ensure that students are reading academic prose so that they encounter academic vocabulary. They also need to read within their subject areas whenever and wherever possible to ensure they encounter bundles that are specific to their chosen discipline. Reading widely and extensively should be supported by direct instruction on lexical bundles.

## WHAT CAN TEACHERS DO?

When deciding what to do with specialised bundles to understand one text versus bundles for wider learning, teachers need to keep several key points in mind. First of all, focus on learning and teaching lexical items today that will be useful for learners tomorrow (Nation, 2009). That is, think carefully about the purposes of learners. Another important point is to be aware that vocabulary knowledge builds incrementally (Schmitt, 2000), and the multiple focused encounters in context and in classrooms should help build this knowledge. Leaving learning to chance encounters with lexical bundles in texts is not a reliable way to build knowledge. Low frequency items in texts risk being overlooked by learners, particularly if they occur in less prominent positions in texts (Coxhead, 2008b). A principled approach such as Nation's (2008) four strands of (a) meaning-focused input, (b) meaning-focused output, (c) language-focused learning, and (d) fluency may provide a structure for a language curriculum that ensures balance of time and effort.

Teachers need to decide on principles which lexical bundles to teach. Principles such as frequency, range, teachability/learnability, and how useful the bundles can support decision making (see Gouverneur, 2008; Nation, 2001). It is also important to be aware that learners may be resistant to learning two words or phrases when learning one word alone may seem hard enough (Coxhead, 2008a). Using words and phrases in writing can be difficult as learners can struggle to combine or match the lexical bundles with the context of their own writing and speaking (Coxhead, 2008b). If teachers decide that using the bundles in output is important for learners, teachers should discuss or explain expectations of use in class assignments and assessments (Coxhead, 2008b).

Teachers can draw attention to bundles in class readings/class materials (Nation, 2008) and use them as the basis for some explicit instruction (Kennedy, 2008). Keeping track of bundles that have been subject to attention in class is also crucial, for example through class vocabulary boxes (see Coxhead, 2004), vocabulary notebooks (see Nation, 2001; Schmitt, 2000), or space on a whiteboard. These bundles should be revisited regularly to increase the likelihood of remembering them and to create opportunities for feedback (Webb, 2007). Learners' awareness of the value of lexical bundles for fluency in all four skills (see Nation, 2001; Wray, 2002; Wray & Fitzpatrick, 2008) and for processing and interacting in their student group (Wray, 2002) needs to be raised. Developing students' understanding of the value of deliberate learning on encountering bundles in reading and listening is also important (Nation, 2008 p.122). Learners may have their own sense of words and phrases that are useful to learn for their studies. Their own language can be heavily influenced by their own reading in a subject area and current affairs, as well as everyday language encounters (Coxhead, 2008b).

Concordancing programmes, such as those available on Tom Cobb's website, the Compleat Lexical Tutor (Cobb, n.d., available at http://www.lextutor.ca/), can help teachers and learners create concordances of lexical bundles using electronic versions of classroom texts. Electronic texts can be created by scanning, retyping sections, downloading from the internet perhaps, or through publishers' websites for some textbooks (see Cobb, 1997; Hirsh & Coxhead, 2009; Thurstun & Candlin, 1998). Teachers can use learners' texts (see Gilquin, Granger, & Paquot, 2007; Meunier, 2002) and/or publicly available corpora (see Schmitt, 2000 for lists of corpora and Cobb (n.d.) for access to web-based text analysis tools as well as corpora). Teachers can use concordances from these texts for intensive study or project work in class or for the basis of independent study through activities such as word cards (see Coxhead, 2004; Nation, 2008, 2001). It is important to note that some students may not respond well to concordancing. Kennedy (2008, p. 38) cautions against returning to language teaching as "applied systems" and warns of difficulties in capturing and retaining student motivation with this technique. See also Wible (2008) for more on the "digital turn" for learning multiword expressions.

## FURTHER RESEARCH

Many theoretical and practical tasks are required to make it possible for teachers to provide access for EAP students to the ways in which preformulated and formulaic language underlies academic communication. We need, for example, to know how lexical bundles fit into other patterning. Focusing just on lexical bundles would not give students access to the full range of formulaic, multiword units that are regularly used in academic writing. Applied corpus linguists should concentrate attention on attaining agreement on terminology as much as possible. Nation (2008, p.117), for example, proposes the term "multiword units" rather than "collocations" because the term "collocations" has many meanings to different researchers and teachers. Research reports ought to give teachers more information about how formulaic or multiword patterns are used in context. Additionally, multiword sets could be broken down to show how shorter units combine to make larger ones.

We need to consider the needs of teachers and learners as much as possible in research reports. Further investigation, evaluation, and reporting on classroom approaches to instruction based on lexical bundles will benefit teachers and learners also. Teachers may well already be employing ways to work with lexical bundles that have yet to be reported in the literature. Language teacher education programmes should include instruction on what Sinclair called the Idiom Principle (Erman & Warren, 2000) to help teachers see repeated patterning in language in use and its worth. We also need more publically available corpora like MICASE and more publically available corpus tools such as those on the Compleat Lexical Tutor (Cobb, n.d.) so that teachers (and researchers without much or any funding for their work) can participate in analysis of the linguistic features of language in use.

Finally, we need more study of formulaic language and multiword units in both speech and writing across the academy, including comparisons to the speaking and writing characteristic of particular disciplines. While useful research can result from analysis of transcripts of academic speaking (e.g., Erman, 2007), little analysis has been done of prosodic features of academic speech (Pickering & Byrd, 2008). This newer line of research should result in clearer understanding of the ways in which lexical patterns support academic communication and are linked together in fluent communication.

## THE AUTHORS

Pat Byrd is Professor Emerita in the Department of Applied Linguistics and English as a Second Language at Georgia State University in Atlanta, Georgia. Her work focusses on academic language, especially the relationship between grammar and vocabulary in academic writing, and teaching EAP grammar, vocabulary, and writing. She has written and edited a variety of textbooks, including the English for Academic Success series edited with Joy Reid and Cynthia Schuemann.

Averil Coxhead is a senior lecturer in Applied Linguistics in the School of Linguistics and Applied Language Studies, Victoria University of Wellington. Averil developed and evaluated the Academic Word List (AWL) and is the author of *Essentials of teaching academic vocabulary* (2006). Her current research projects include vocabulary size measurements, vocabulary teaching and learning in secondary schools, and the collocations and phraseology of the AWL in written texts.

## REFERENCES

Altenberg, B. (1998). On the phraseology of spoken English: The evidence of recurrent word-combinations. In A. H. Cowie (Ed.), *Phraseology: Theory, analysis, and applications* (pp.101-122). Oxford: Clarendon Press.

Baker, P. (2006). *Using corpora in discourse analysis*. London: Continuum.

Bauer, L. (1993). *Manual of information to accompany the Wellington Corpus of Written New Zealand English*. Wellington, NZ: Victoria University of Wellington.

Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.

Biber, D. (2006). *University language*. Amsterdam: John Benjamins.

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26(3), 263-286.

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 lune: A festschrift for Geoffrey Leech* (pp.71-93). 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(3), 371-405.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow, England: Pearson Education Limited.

Campbell, N. & Reece, J. (2005). *Biology*. San Francisco: Pearson Benjamin Cummings.

Cobb, T. (1997). Is there any measurable learning from hands-on concordancing? *System*, 25(3), 301-315.

Cobb, T. (n.d.). Compleat lexical tutor. Available at http://www.lextutor.ca/ on March 11, 2010.

Conrad, S. (2008). Writing Myth 6: Corpus-based research is too complicated to be useful for writing teachers. In J. Reid (Ed.), *Myths about teaching writing* (pp.115-139). Ann Arbor: University of Michigan Press.

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes, 23*, 397-423.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(2), 213-238.

Coxhead, A. (2004). Using a class vocabulary box: How, why, when, where, and who. *RELC Guidelines*, 26(2): 19–23.

Coxhead, A. (2006). *Essentials of teaching college vocabulary.* Boston: Houghton Mifflin.

Coxhead, A. (2008a). Phraseology and English for academic purposes: Challenges and opportunities. In F. Meunier & S. Granger (Eds.), *Phraseology in language learning and teaching* (pp.149-161). Amsterdam. John Benjamins.

Coxhead, A. (2008b). *Using vocabulary from input texts in writing tasks*. Unpublished PhD thesis. Wellington: School of Linguistics and Applied Language Studies, Victoria University of Wellington.

Erman, B. (2007). Cognitive processes as evidence of the idiom principle. *International Journal of Corpus Linguistics*, 12(1), 25-53.

Erman, B., & Warren, B. (2000). The idiom principle and the open-choice principle. *Text*, 20, 29-62.

Folse, K. (2004). *Vocabulary myths: Applying second language research to classroom teaching*. Ann Arbor, MI: University of Michigan Press.

Gilquin, G., Granger, S., & Paquot, M. (2007). Learner corpora: The missing link in EAP pedagogy. *English for Academic Purposes*, 6, 319-335.

Gouverneur, C. (2008). The phraseological patterns of high-frequency verbs in advanced English for general purposes: A corpus-driven approach to EFL textbook analysis. In F. Meunier & S. Granger (Eds.), *Phraseology in foreign language learning and teaching* (pp.223-243). Amsterdam: John Benjamins.

Granger, S. & Meunier, F. (2008). Phraseology in language learning and teaching: Where to from here? In F. Meunier & S. Granger (Eds.), *Phraseology in foreign language learning and teaching* (pp.247-252). Amsterdam: John Benjamins.

Heatley, A., Nation, P. & Coxhead, A. (2002). Range. Available from http://www.victoria.ac.nz/lals/staff/paul-nation/nation. aspx on 11 March 2010.

Hirsh, D. & Coxhead, A. (2009). Ten ways of focusing on science-specific vocabulary in EAP. *English Australia Journal*, 25(1), 5-16.

Hunston, S. & Thompson, G. (2000). Evaluation in text: Authorial stance and the construction of discourse. Oxford: Oxford University Press.

Hyland, K. (1998). *Hedging in scientific research articles*. Amsterdam: John Benjamins.

Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, 27(1), 4-21.

Hyland, K. & Tse, P. (2007). Is there an "academic vocabulary"? *TESOL Quarterly*, 41(2), 235-253.

Johannson, S. (1978). *Manual of information to accompany the Lancaster/Oslo/Bergen Corpus of British English, for use with digital computers*. Oslo: Department of English, University of Oslo.

Jones, M. & Haywood, S. (2004). Facilitating the acquisition of formulaic sentences: An exploratory study in an EAP context. In

N. Schmitt (Ed.), *Formulaic sequences* (pp.269-291). Amsterdam: John Benjamins.

Kennedy, G. (2008). Phraseology and language pedagogy: Semantic preference associated with English verbs in the British National Corpus. In F. Meunier & S. Granger (Eds.), *Phraseology in foreign language learning and teaching* (pp.21-31). Amsterdam: John Benjamins.

Kucera, H. & Francis, W. N. (1967). *A Computational analysis of present day American English*. Providence, Rhode Island: Brown University Press.

Meunier, F. (2002). The pedagogical value of native and learner corpora in EFL grammar teaching. In S. Granger, J. Hung & S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp.119-142). Amsterdam: Benjamins.

Nation, P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Nation, P. (2008). *Teaching vocabulary: Strategies and techniques*. Boston: Heinle, Cengage Learning.

Nation, P. (2009). *Teaching ESL/EFL reading and writing*. New York: Routledge.

Pawley, A. & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (pp.191-226). new York: Longman.

Pickering, L. & Byrd, P. (2008). Investigating connections between spoken and written academic English: Lexical bundles in the AWL and in MICASE. In D. Belcher & A. Hirvela (Eds.), *The Oral/Literate Connection: Perspectives on L2 speaking, writing and other media interactions* (pp.110-132). Ann Arbor: University of Michigan Press.

Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge: Cambridge University Press.

Scott, M. (2006). WordSmith Tools 4. Available at www.lexically.net/wordsmith on March 11, 2010.

Sinclair, J. & Mauranen, A. (2006). *Linear unit grammar*. Amsterdam: John Benjamins.

Stubbs, M. & Barth, I. (2003). Using recurrent phrases as text-type discriminators: A quantitative method and some findings. *Functions of Language*, 10(1), 61-104.

Tarone, E. S. Dwyer, S. Gillette, S., & Icke, V. (1981). On the use of the passive in two astrophysics journal papers. *The ESP Journal*, 1(2), 123-140.

Thurstun, J., & Candlin, C. N. (1998). Concordancing and the teaching of the vocabulary of academic English. *English for Specific Purposes*, 17(3), 267-280.

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, 28(1), 46-65.

Wible, D. (2008). Multiword expressions and the digital turn. In F. Meunier & S. Granger (Eds.), *Phraseology in foreign language learning and teaching* (pp.163-181). Amsterdam: John Benjamins.

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.

Wray, A. (2008). *Formulaic language: Pushing the boundaries*. Oxford: Oxford University Press.

Wray, A., & Fitzpatrick, T. (2008). Why can't you just leave it alone? Deviations from memorized language as a gauge of nativelike competence. In F. Meunier & S. Granger (Eds.), *Phraseology in foreign language learning and teaching* (pp.123-147). Amsterdam: John Benjamins.