



This is the author's version of a work that was accepted for publication in the following source:

Eriksson, Andreas. 2012. 'Pedagogical perspectives on bundles: Teaching bundles to doctoral students of biochemistry'. In James Thomas & Alex Boulton (eds). *Input, Process and Product: Developments in Teaching and Language Corpora*. Brno: Masaryk University Press, 195-211.

**Notice:** Changes introduced as a result of publishing processes such as copy-editing and formatting may not be reflected in this document. For a definitive version of this work, please refer to the published source.

Andreas Eriksson  
Chalmers University  
of Technology

# Pedagogical perspectives on bundles: Teaching bundles to doctoral students of biochemistry

This paper explores lexical bundles from a pedagogical perspective, targeting two issues in particular: the selection of bundles for teaching purposes and students' perception of different learning activities. Two small corpora of doctoral student writing and published articles were compiled for use in two workshops dedicated to lexical bundles. The assessment of learning activities used in the workshops was based on classroom observations, student notes and discussions. The results underline the importance of taking into account the field in which the students work when selecting bundles. They also indicate that the most useful activities are those in which the students started out by speculating about usage in their own discipline and those in which they were asked to use bundles within the context of their own research. The implication is that these activities offered opportunities for deep level processing.

## 1. Introduction

In recent years, notable steps have been taken to add new perspectives to the application of corpora for teaching purposes. For instance, Charles (2007) has addressed the potential of combining top-down and bottom-up approaches to academic writing by means of corpora, O'Sullivan (2007) and Flowerdew (2008, 2009) have discussed the scaffolding of learning activities involving corpora, and Granger and Meunier (2008) have tried to strengthen the links between second language acquisition and corpus pedagogy by gathering researchers from several fields of linguistics in a volume on the teaching and learning of phraseology. These studies are indicative of important developments going on within the field, but there are still areas and perspectives that remain relatively uncharted.

One area where the pedagogical implications remain fairly unexplored is the field of frequently occurring word combinations, also known as clusters, chunks or lexical bundles. Instead, studies in this area have primarily had a descriptive focus. Examples of studies which have provided detailed and helpful descriptions of the use of bundles in different genres are Biber (2006), Cortes (2004), Hyland (2008a, 2008b), Scott and Tribble (2006) and Simpson-Vlach and Ellis (2010). When pedagogical implications are mentioned, it is primarily through general comments indicating what bundles may be of importance for foreign or second language learners of English (e.g. Cortes 2004; Hyland 2008b; Simpson-Vlach & Ellis 2010) or through general discussions of pedagogical implications (Byrd & Coxhead 2010; O'Keeffe et al. 2007). The only major studies that have directly addressed the teaching of bundles in classroom settings are Jones and Haywood (2004) and Cortes (2006). All in all, this means that comparatively little attention has been given to what bundles to teach and how to teach them, despite the fact that it has been argued that the use of field-specific bundles signals adherence to a particular genre, and that bundles therefore are important for people who want to write professionally within that genre (Hyland 2008a: 5, 2008b: 42; Wray 2002).

The present paper addresses bundles from a pedagogical perspective. The aim is to raise pedagogical concerns which may inform the development of activities and teaching practices to foster the learning of bundles. The study consists of two main parts: the selection and identification of bundles for teaching purposes, and the investigation of students' work on and perception of a number of learning activities using bundles. The participants in the study are doctoral students of biochemistry and biotechnology, all non-native speakers of English. The students were invited to participate in two three-hour workshops on the use of bundles (the term 'multi-word units' was used in the workshops) in academic writing in their discipline. In the workshops, the students used texts written by themselves and contrasted these with a small corpus consisting of published scientific articles from their discipline.

Following Biber (2006: 134), bundles are defined here as "simply the most frequently occurring sequences of words." Biber then develops this definition by adding a minimum frequency cut-off point and a minimum number of texts across which bundles must be distributed. As he points out, the frequency cut-off is somewhat arbitrary, but two common limits are 40 per million words (Biber 2006) or, for some studies on four-word bundles, 20 per million words (Cortes 2006; Hyland 2008a).

## 2. Applying corpora to the teaching and learning of lexical bundles

This chapter introduces issues in corpus pedagogy that directly affect the present study. Section 2.1 reviews the application of small-scale corpora and student-compiled corpora in teaching, Section 2.2 presents studies dealing with the teaching of bundles, and Section 2.3 accounts for studies on vocabulary teaching and learning.

### 2.1. Applying small-scale and student corpora in teaching

The two corpora used in this study are small, in-house corpora (see Section 3.1). As noted by Flowerdew (2009: 396, 2010: 444), it has become quite common to build specialised corpora for teaching purposes within the fields of ESP and EAP, and several authors have reported on the pedagogical effectiveness of such corpora (e.g. Charles 2007; Flowerdew 2004; Tribble 2001).

One particular type of small-scale corpus consists of material produced by a single student. Such corpora can be used to compare for example student writing with expert writing. The idea of personalised corpora for such comparisons was launched by Turnbull and Burstun (1998), but probably the best-known example of this type of approach is Lee and Swales (2006), who designed a corpus-based EAP course for doctoral students (see also Aston 2002; Ngouchi 2004; Seidlhofer 2000). In this course, the students compiled mini-corpora of their own writing and of published papers in their disciplines. Lee and Swales (2006: 71) noted that one of the main advantages of using material close to the students' disciplines was that it seemed to increase engagement in the tasks they performed. It seems likely that this motivational factor is at least partly linked to the fact that personalised corpora make it possible to individualise teaching (Mukherjee & Rohrbach 2006: 220; Flowerdew 2009: 411).

The overall idea of having doctoral students contrast their own writing with published articles in their discipline is employed in this study as well, but with a narrower scope than in Lee and Swales (2006) as the focus is on lexical bundles. In addition, the students in the present study did not compile any corpora themselves but only submitted their own texts to the teacher, who then converted the texts into a format that could be used with the software AntConc 3.2.1 (Anthony 2007).

### 2.2. Previous studies on the teaching of bundles

Jones and Haywood's (2004) and Cortes' (2006) pioneering, exploratory studies on the teaching of bundles share some common ground and perspectives, the attempt at measuring the effects of the teaching of bundles in learner production being just one. At the same time, they cover different ends of a spectrum as they deal with completely different learner groups. Jones and Haywood worked with non-native speakers of English intending to study at British universities, while Cortes' students were native speakers of English studying history in their third or fourth year.

Jones and Haywood (2004) introduced lexical bundles in a so-called pre-sessional EAP course for non-native speakers from different disciplines. Considering the student profile, they

selected bundles that were frequent and common in many disciplines as the target bundles of the course (Jones & Haywood 2004: 274). The study aimed to raise students' awareness of lexical bundles, to improve their use of lexical bundles, and to try to find strategies that would facilitate learning them. The teaching featured a range of activities, such as the reading of texts where certain bundles had been highlighted, the classification of bundles according to function, and the analyses of concordance lines and text extracts.

Student development was assessed through several pre- and post-tests. The results of these tests indicated some improvement in both awareness and production, but the results could not be substantiated through statistical tests as the numbers were too low. Jones and Haywood (2004: 290) see time as a limiting factor, both in the sense that comparatively little time was spent on bundles during the ten weeks of the study as well as in the sense that ten weeks may have been too short a period to have significant effects on student awareness and production.

In contrast to Jones and Haywood, Cortes (2006) worked with native speakers of English in a ten-week, writing-intensive history class. She compared eight students' use of four-word bundles in writing with that of published scientific articles in the same field. The teaching of bundles in the course consisted of five 20-minute sessions focusing on bundles. For these sessions, 35 four-word bundles were selected from an 800,000-word corpus. The sessions involved analyses of functions, exercises of fill-in-the-blanks and multiple choice type, corrections of inappropriate usage, and paraphrasing tasks.

Like Jones and Haywood, Cortes was interested in assessing student development and did so via writing assignments handed in during the course. Three assignments were collected at various stages of the course: the first one before the first of the five micro-lessons, the second one after the third lesson, and the third one at the end of the course. In all of these assignments, the frequencies of all target bundles were recorded.

The results of Cortes' study show that the target bundles were found relatively infrequently in the student texts, and it was not possible to see any development insofar as the students did not use more bundles in the texts written after the micro-lessons than in the texts written beforehand. The results are thus similar to Jones and Haywood's in that a ten-week period seems to be too short a period for finding substantial development in students' use of bundles. In Cortes' case, it must also be questioned whether it is reasonable to expect a clear growth in terms of frequency in the texts she collected – and even whether such a development is desirable. For instance, there are comparatively few opportunities for a learner to actually use a particular bundle in one single text, as most bundles are often used only once or twice even in published texts. The chances for an individual student to identify a passage where the use of a particular bundle is suitable and where that student also chooses a bundle to express what he/she wants to express are therefore comparatively small. It is thus not very surprising if it is not possible to show a significant increase in the use of bundles as a result of a short-term teaching intervention when using student texts to assess development.

Both these studies used a variety of activities but comment very sparingly on the effectiveness of different exercises. Cortes (2006: 401) speculates that one of the reasons she did not find any increase in her students' use of bundles may have been that the in-class activities did not make students more independent in their use of bundles in writing, but she does not discuss what types of exercises may have supported such a development. She does say, however, that in informal

interviews, students claimed that the lessons had made them more aware of functions performed by bundles. Jones and Haywood (2004: 277) attempted to design exercises on bundles that would generate deep processing; they also describe sequences where students were prompted to notice grammatical as well as semantic preferences of particular bundles. However, the effectiveness of the exercises is only commented on indirectly as the authors acknowledge the importance of research about how to support students' learning of phraseology.

The most recent study that addresses the teaching and learning of bundles is Byrd and Coxhead's (2010) study of lexical bundles in academic writing. They use the corpus that was used for the development of the Academic Word List (Coxhead 2000) to identify bundles that are used across the four subject areas of the corpus: law, commerce, arts and science. They guide readers through their process of identifying lexical bundles frequently found in all four registers. In the end, 21 widely used bundles are selected, and these bundles are then used in discussions of, for example, structural characteristics and discourse functions. The authors also discuss six challenges for teachers who are to teach lexical bundles in academic writing and/or more generally in EAP. Some of the challenges they discuss are how to apply frequency-based lists of lexical bundles in teaching, how to deal with four- or five-word bundles that hold shorter bundles within them, and how to deal with the lack of information about the context in which specific bundles can be used. Byrd and Coxhead end their article by asking for more studies based on classroom experience of the teaching of bundles.

A further study that touches on the teaching of bundles is O'Keeffe et al.'s (2007: 58-79) concise overview of multi-word expressions or chunks. Their focus is on spoken language, and they point to exercises and strategies that may be useful when learning chunks, such as inserting and highlighting chunks in texts and dialogues. They do not, however, go into great detail about teaching, which, coupled with their focus on spoken language, renders their discussion less relevant here.

### 2.3. Approaching the teaching and learning of bundles

As pointed out by Jones and Haywood (2004: 271), little has happened in the area of phraseology research that has influenced classroom practice. As a result, studies into the teaching of phraseology often build on general vocabulary teaching methodology. Consequently, like Jones and Haywood (2004) and Coxhead (2008), the design of activities for the student workshops were primarily influenced by Nation's (2001) three psychological processes that can lead to words being remembered: noticing, retrieving and generating (cf. also Lewis' [2000] observe-hypothesise-experiment paradigm).

The first of Nation's (2001) three processes for successful vocabulary learning is noticing, which quite simply means that the learner pays attention to a word or phrase. Noticing can take many different forms, from looking up a word in a dictionary to negotiating its meaning. The second process is retrieval, which refers to form-meaning mappings that are made when receiving or producing language; as it can only occur after a word has been noticed, it is thus closely linked to repetition. The effectiveness of repetition is dependent on the length of time between previous encounters with a word. The final process that is important for vocabulary learning is generating, whereby a word is used in a way that is different from the way it was used in the learner's previous

encounter(s) with it. Like retrieving, generating can be either productive or receptive, but there are different degrees of generation depending on factors such as the similarity between cotexts and/or contexts.

Because of the somewhat limited availability of theoretical underpinnings to base bundle activities on, it is worth looking at approaches that have been used and found to be effective in other areas of learning research. One promising approach is to combine Nation's above-mentioned processes and levels of student activity with the SOLO taxonomy, i.e. Biggs and Collins' (1982) structure of observed learning outcomes. In brief, the SOLO taxonomy outlines two principal stages in learning development: a quantitative and a qualitative stage. This taxonomy suggests that students move from a shallow approach to learning and understanding to a deeper one as their knowledge moves from being primarily quantitative to including more qualitative aspects. Biggs and Tang (2007: 80) exemplify how the taxonomy can be used to outline intended learning outcomes at different levels and also how these can be connected to student activities at each level. From the perspective of traditional vocabulary learning exercises, this would mean that exercises where students memorise, identify and imitate, such as fill-in-the-blanks or underlining exercises, are less complex and tend towards the quantitative end of the spectrum, whereas exercises where students analyse examples and use words in the context of their own research are more complex and therefore more qualitative in character.

## 3. The selection of bundles

As noted by Jones and Haywood (2004: 274), the question of what bundles to select for teaching purposes is not straightforward. The most common approach to date has been to focus on highly frequent bundles (Cortes 2006; Hyland 2008a; Jones & Haywood 2004). This strategy seems to be well-motivated in some situations, for instance when working with students who have comparatively little experience of academic writing. However, it is more questionable in the context of the present study, as the focus may then be on bundles that students already know how to use, perhaps primarily as a result of being socialised into the discourse of a particular discipline and of having encountered many of these frequent bundles when reading texts within that discipline. In order to avoid working with bundles that the students already knew, the focus here was therefore on bundles found in the published articles but not in the students' texts.

### 3.1. Corpora

Two small corpora were compiled for the present study. The first consisted of 25 articles within the fields of biochemistry and biotechnology. Five articles each were taken from the journals *Biomaterials*, *Bioresource and Technology*, *International Journal of Pharmaceutics*, *Journal of Colloids and Interface*, and *Vaccine*. The articles were selected from these journals because the PhD students participating in the study regarded them as central to their own research. The 25 articles generated a corpus of approximately 105,000 words, here referred to as the Biotech corpus. Abstracts, table and figure captions and reference lists were not included in the corpus.

The second corpus consisted of texts written by six doctoral students, five of whom participated in at least one of the bundle workshops. The participants submitted between two and six texts each. The PhD texts may not be fully comparable as the students were at different stages of their studies, and some of the articles were written by several authors (the students participating in the present study being the first author of the article in all such cases), but the texts represent the type of writing that these doctoral students do. The corpus, henceforth called the PhD corpus, consisted of approximately 100,000 words.

### 3.2. Criteria and procedure for the selection of bundles

The selection of bundles was based on a fairly crude measure: bundles that were found in the published texts but not in the PhD texts were considered interesting. Consequently, the selection always involved a comparison of frequencies between the Biotech and PhD corpora. Because the corpora were fairly similar in size (105,000 vs. 100,000 words), the comparisons were based on raw frequencies. However, not all bundles where there was a difference in frequency between the two corpora were necessarily interesting. In order to avoid selecting bundles that (1) were found only rarely in the Biotech corpus, (2) already seemed to be used by the doctoral students, and (3) were used in few of the Biotech texts, a number of simple criteria were established. In order for a bundle to be selected as interesting for the PhD students of the study, it had to:

- occur at least eight times in the Biotech corpus<sup>1</sup>;
- occur less than six times in the PhD corpus;
- be used in at least three of the articles in the Biotech corpus;
- have a frequency difference of at least five tokens between the Biotech and PhD corpora.

In accordance with these criteria, the bundle *the use of* with a frequency of 37 and 17 in the Biotech and PhD corpora respectively, was not included in the study, as it appeared more than five times in the PhD corpus. In contrast, to *determine whether* was included as there were nine examples of it in four texts in the Biotech corpus but only one example in the PhD corpus.

The criteria were obviously fairly crude and there is a risk of missing bundles that could be important (at least for some learners) and of including bundles that are not particularly important. The reason for using this fairly crude methodology was that it was an attempt at developing a rough-and-ready approach for teachers who want to target bundles that may be worth teaching in an academic writing class.

In order to shed some light on advantages and disadvantages of different selection strategies, three already published lists of bundles were also brought into the selection process. The lists were Hyland's (2008a) lists of the 50 most frequent four-word bundles in the fields of engineering and technology, and Simpson-Vlach and Ellis' (2010) Academic Formulas List (AFL) of which only the generic and written parts of the AFL were used. Hyland's lists are based on corpora from two disciplines that are relatively close to biochemistry and biotechnology: biology and engineering.

The Academic Formulas List is more general in nature, covering bundles that are important in general academic language. The three lists were primarily used in order to check to what extent bundles that were selected by means of the comparison between the Biotech and PhD lists could also be found on Hyland's lists and the AFL. The results of such a comparison should give some indication of how important the issue of disciplinarity is for bundles in academic writing.

In addition, Hyland's lists and the AFL were given a small role in the identification of bundles used in the Biotech corpus but not in the PhD corpus. Because the Biotech and PhD corpora are comparatively small, and because Hyland's lists and the AFL represent more established lists of academic bundles, the cut-off points for the frequency in the Biotech corpus was set slightly lower (six tokens rather than eight). Consequently, bundles with a frequency of six or seven in the Biotech corpus were also included in the list of bundles worth teaching, provided that they were on one of Hyland's lists or the AFL and that there was a difference of at least five tokens between the Biotech and PhD corpora.

Three- to five-word bundles were extracted by means of the cluster and n-gram function in AntConc (Anthony 2010). The bundles that met the frequency criteria were then checked via the concordance function to make sure that only bundles that were used in at least three articles were included in the study.

### 3.3. Results of the selection of bundles

Before going into the selection of bundles, it is worth mentioning one finding about the distribution of bundles in the Biotech and PhD corpora. As shown by Hyland in his comparison of bundles in published articles and PhD and Master's theses, students' use of bundles is not simply a matter of students using fewer bundles than more experienced writers in the discipline. On the contrary, Hyland (2008b: 50) showed that bundles were more frequent in the thesis genres than in the published articles. Similar results were found in the present study when searching for all the written and generic bundles from the AFL in the Biotech and PhD corpora. There were 1,173 tokens of these bundles in the PhD corpus, but only 963 in the Biotech corpus. Hyland gives two main reasons for the differences found in his study. First of all, he argues that there are genre differences between the thesis genres and the article genre, but that explanation seems less valid in the present study as texts belong to the same genre. Instead, it seems more likely that Hyland's second point offers more of an explanation here. Hyland argues that differences can be due to less experienced writers being more dependent on prefabricated chunks than experienced writers. It also seems as if PhD students tend to rely on some bundles that they use frequently. These findings are important as they show that the learning task that these learners are facing is not simply a matter of frequency.

The first step in the selection of bundles was to produce frequency lists of bundles in the Biotech and PhD corpora. Table 1 shows the most frequent 3- to 5-word bundles in the two corpora. The bundles that appear on both lists have been italicised.

The columns show that there is a substantial overlap between the bundles found on the two lists despite some clear differences in terms of frequency, for instance between *the presence of* and *in order to*.

<sup>1</sup> The figures have not been normed per million words but a frequency cut-off of 8 should meet the commonly used cut-offs of 20 or 40 per million words, considering that the corpus consists of 105,000 words.



Table 1. The most frequent 3- to 5-word bundles in the Biotech and PhD corpora

Biotech	frequency	PhD	frequency
<i>the presence of</i>	60	<i>in order to</i>	87
<i>due to the</i>	49	<i>the amount of</i>	65
<i>the use of</i>	37	<i>the degree of</i>	47
<i>the concentration of</i>	34	<i>the effect of</i>	47
<i>was used to</i>	34	<i>the number of</i>	43
<i>as well as</i>	33	<i>as well as</i>	42
<i>shown in fig.</i>	31	<i>compared to the</i>	38
<i>in order to</i>	29	<i>the concentration of</i>	35
<i>in the presence</i>	26	<i>in this study</i>	34
<i>the effect of</i>	26	<i>the yield of</i>	32
<i>was added to</i>	25	<i>due to the</i>	30
<i>in this study</i>	22	<i>effect on the</i>	27
<i>the absence of</i>	22	<i>one of the</i>	25
<i>in the presence of</i>	21	<i>the formation of</i>	25
<i>no significant differences</i>	21	<i>related to the</i>	24
<i>to determine the</i>	21	<i>the influence of</i>	24
<i>a function of</i>	20	<i>the presence of</i>	24
<i>compared to the</i>	20	<i>be explained by</i>	22
<i>detected in the</i>	20	<i>was added to</i>	21
		<i>used in this</i>	21
		<i>added to the</i>	21

Table 2. Bundles selected for teaching purposes on the basis of comparisons between the Biotech and PhD corpora

Bundle	Biotech (frequency)	PhD (frequency)	Hyland	AFL
<i>was used to</i>	34	4		
<i>the absence of</i>	22	3	(in the absence of)	(in the absence of)
<i>no significant differences</i>	21	1		
<i>detected in the</i>	20	2		
<i>in the absence</i>	19	1	(in the absence of)	(in the absence of)
<i>in the absence of</i>	19	1	Yes	Yes
<i>in the present</i>	19	4	(in the present study)	(in the present study)
<i>the addition of</i>	19	5		
<i>significant differences in</i>	18	0		
<i>any of the</i>	16	1		
<i>in addition to</i>	15	5	(in addition to the)	

Bundle	Biotech (frequency)	PhD (frequency)	Hyland	AFL
<i>the concentrations of</i>	14	0		
<i>were detected in</i>	14	0		
<i>results suggest that</i>	13	1		
<i>the cost of</i>	13	1		
<i>applied to the</i>	12	2		
<i>the end of the</i>	11	4	Yes	
<i>were allowed to</i>	11	1		
<i>consistent with the</i>	10	0		
<i>it was found that</i>	10	3	Yes	
<i>not significantly different</i>	10	0		
<i>the nature of</i>	10	1		Yes
<i>these results suggest that</i>	10	1		
<i>we found that</i>	11	2		
<i>as previously described</i>	9	0		
<i>it should be</i>	9	3		(it should be noted)
<i>the magnitude of</i>	9	0		
<i>magnitude of the</i>	9	0		
<i>the magnitude of the</i>	9	0		
<i>one or more</i>	9	0		
<i>relative to the</i>	9	0		
<i>to determine whether</i>	9	0	Yes	
<i>be attributed to</i>	8	1		
<i>cost of the</i>	8	1		
<i>have been shown to</i>	8	3		(been shown to)
<i>listed in table</i>	8	0		
<i>the course of</i>	8	0		
<i>the feasibility of</i>	8	0		
<i>the surface area</i>	8	0		
<i>to the presence</i>	8	0		(the presence of)
<i>were determined by</i>	8	0		(is determined by)
<i>have shown that</i>	7	2		Yes
<i>in conjunction with</i>	7	0		Yes
<i>it is possible that</i>	7	1	Yes	
<i>the presence of the</i>	7	2	Yes	
<i>may be due to</i>	6	1	Yes	
<i>a high degree</i>	6	0		Yes
<i>appears to be</i>	6	1		Yes
<i>in response to</i>	6	0		Yes
<i>the change in</i>	6	1		Yes

More careful comparisons show that the PhD students use the most frequent bundles on the Biotech list. In fact, only four of the bundles on the list of the most frequent bundles in the Biotech corpus were found less than six times in the PhD corpus (*was used to, the absence of, no significant differences and detected in the*). All in all, these findings seem to indicate that lexical bundles are learnt without explicitly being taught, and that this learning seems to be incidental and implicit as these students had not previously been taught lexical bundles. It is not unexpected that some of the most frequent bundles are acquired implicitly (Ellis 2008: 7-8), as these are the ones that learners are likely to encounter quite often when reading articles within their own discipline. These findings ultimately lend support to the decision not to rely solely on frequency when selecting bundles to be used in the workshops of the present study. However, it is important to note that these findings must be viewed in light of the fact that the students are doctoral students. It is thus more likely that they will have read extensively within their disciplines and thus have had a chance to encounter the most frequent bundles fairly often. The extent to which the principles for selecting bundles for teaching purposes can rely on students' previous encounters with particular bundles is thus partly dependent on student experience of a particular field or genre. Byrd and Coxhead (2010: 56) make an important point when claiming that it is a risky strategy to base the learning of bundles purely on students' potential encounter with bundles, as less frequent items might be missed.

Turning next to the actual selection of bundles, the comparison of bundles in the Biotech and PhD corpora rendered the 50 bundles shown in Table 2. The two middle columns show the raw frequencies of each bundle in the Biotech and PhD corpora respectively. The fourth column shows whether the bundle can be found on Hyland's (2008b) lists or on the AFL (Simpson-Vlach & Ellis 2010). Bundles found in exactly the same form on one or the other list are marked 'yes'; bundles found in a similar form are presented in parentheses; blank spaces show that bundles do not appear on the list. Finally, it is worth noting that bundles with a frequency of 6 or 7 in the Biotech corpus were included provided that they could be found on Hyland's lists or the AFL.

As can be seen, more than half of the bundles identified by means of a discipline-specific list would not have been found if only Hyland's list and the AFL had been used. This shows that discipline-specific comparisons can make a considerable difference in the selection process: there is an obvious risk of missing important bundles if the factor of discipline is not taken into consideration.

#### 4. Workshops: Participants and design of activities

The five doctoral students who participated in the workshops had little or no experience of the use of corpora and corpus software. In terms of experience of analysing academic writing, four participants had formal training in this through an academic writing course for doctoral students. The principal aim of the course was to raise students' awareness and understanding of the structure and rhetoric of research articles in general and of disciplinary writing in particular. The students who had taken this course thus had some experience of analysing academic writing but the course did not involve any work with language corpora.

During the workshops, student behaviour was observed, and afterwards a copy was taken of the notes that the students had made. In addition, the participants were asked to evaluate the workshops and each workshop ended with a short discussion of the advantages and disadvantages of the different exercises.

A set of exercises was designed for each workshop featuring four major types of activities:<sup>2</sup>

1. Activities focusing on searches in the Biotech and the PhD corpora (or the students' own subcomponents of the PhD corpus). The major activity involved comparing the use of specific bundles in the two corpora, and trying to draw conclusions about usage and/or functions on the basis of these comparisons; cf. task (c) in Figure 1.
2. Activities employing student usage (or student beliefs about usage) in their discipline as the starting point. Students were sometimes asked to speculate about what words or phrases they used to express certain functions, for instance what wording they used to indicate uncertainty in their results; cf. tasks (a) and (b) in Figure 1.
3. Activities where students were asked to use bundles in the context of their own research. Since the workshops were fairly short, this activity only involved the writing of sentences where the participants saw that a particular bundle could be used; cf. task (d) in Figure 1.
4. Activities focusing on particular functional categories. These functional categories can be presented at various levels of abstraction, two of which were tested in the present study. First of all, one exercise focused on bundles used to express comparison and contrast according to the categorisation of expressions in Simpson-Vlach and Ellis (2010: 499). Secondly, one exercise took Hyland's (2008b) division into research-oriented, text-oriented and participant-oriented bundles as its starting point; cf. task (a) in Figure 1.

As indicated in the list above, the activity types can be illustrated by means of the first exercise from the first workshop shown in Figure 1.

The observations and the student notes made in connection with this exercise showed first of all that the students had a fairly clear idea of what words they used to indicate uncertainty. Some of the examples they gave were *it might be due to, it could be due to* and *the results may indicate that*. In task (c), the students here noticed that there was a clear contrast between their own work and published texts. For instance, many of them realised that the phrase *might be due to* could not be found in the published papers but that they used it quite often themselves. Finally, in task (d), the students were asked to use the bundles in (b) in the context of their own research. They could do so fairly easily and seemed to take the task seriously. The sentences developed were stand-alone sentences but they clearly fitted within the context of the students' research.

<sup>2</sup> The distinction between exercises and activity types was made purely for analytical reasons. Each exercise contained at least one, often several, of the four activity types.

Figure 1. Sample exercise from the first workshop

**Qualifying claims**

a) How do you express / indicate indeterminacy in your claims, and/or how do you avoid making too strong claims in your texts (for instance if you know that your results are not conclusive)?

---



---

b) Which of the following multi-word units do you think are commonly used in texts from your discipline?

- *it looks like*
- *may be due to*
- *appear(s) to be*
- *might be able to*
- *results suggest that*
- *at least in*

c) Open the texts in the folder “Published texts” in AntConc. Search for the multi-word units above in these texts.

d) Try to come up with an example in your own writing, perhaps in an article you are currently working on, where it would be possible to use one or two of the multi-words units common in academic writing listed above.

---



---

The evaluation of the four major types of activities showed that type 2 and type 3 activities were the most useful, the former because the student hypotheses often served as a good background for comparisons between their own writing and published texts. It should be pointed out, however, that not all examples where there was a difference between the students’ hypotheses and actual usage were considered equally useful. For instance, when working with the bundle to *determine whether*, the students realised that, contrary to their expectations, it was used not only in the methodology section but also in the results section. It is indeed difficult to see what the implications of such a finding would be, particularly as the findings did not really contradict their initial hypothesis. It thus seems that it is only in situations where hypotheses are combined with comparisons adding new insights to the students’ own writing that this type of activity becomes truly meaningful.

The third type of activity was successful in the sense that students were able to write sentences that seemed to work in the context of their own research. It also meant that they had to devote some time to actually engaging with how a particular bundle is used, and were prevented from moving on to another bundle until they had thought about the use and meaning of the previous one.

The value of type 4 activities was not as obvious as for types 2 and 3. On the one hand, when working on exercises listing many bundles performing one particular rhetorical function, the

students tended to work very quickly and focus on frequency comparisons rather than detailed analyses of concordance lines and/or the larger context. They thus saw that there are many different ways of setting up a passage of comparisons and contrasts, but did not really learn more about such passages. The results thus indicate that a design containing a list of bundles can easily turn into a frequency counting exercise that does not really help students learn more about a particular function. Such activities would be indicative of what Biggs and Tang (2007: 22) refer to as a surface approach to learning.

At the same time, the students emphasised that they would like to have more exercises on specific functions. This request seems to be connected with a desire to have more alternatives for expressing similar functions, as pointed out in one of the discussions after the workshops. For example, one student stated that the most useful exercise in the second workshop was for presenting results, as it displayed ways of expressing results that she had not used before.

## 5. Teaching lexical bundles: Reflection and future concerns

The aim of the present study was to raise pedagogical issues on the teaching and learning of bundles. It has done so by targeting two issues affecting classroom practice: the selection of bundles and the development of learning activities. The selection issue is important as there is typically a wide range of bundles to choose from, and it is therefore often necessary to make decisions about what bundles to focus on in a teaching situation. The results of this study suggest that disciplinarity and specialisation have to be taken into account when deciding what bundles to consider.

The workshops quite clearly showed that the students generally appreciated exercises where they speculated about usage within their own discipline and where they used bundles within the context of their own research. One reason for the students’ positive evaluation may be that they found it motivating to work with exercises based on their own writing (Lee & Swales 2006: 61). Another reason may be that the combination of student hypotheses and comparisons between the PhD and Biotech corpora leads to more thorough analyses, as students want to find out how they can improve their own writing. The strength of exercises where students produce something in relation to their own research is that they engage in a type of activity that is necessary for vocabulary learning (Nation 2001) and which involves a more advanced level of understanding (Biggs & Tang 2007).

Another potential strength of activities where students have to use bundles within the context of their own research is that they partly address a type of criticism directed at the use of corpora in teaching, namely that the language presented in corpora is decontextualised (Widdowson 2000: 7-8). How such criticism is interpreted partly depends on the definition of context, but if one adopts a fairly broad interpretation, it is true that a great deal of contextual information can be obscured in a corpus-based approach, for instance about the actual circumstances in which the text was produced. By having students integrate bundles into their own writing, the activity is set within a context that they are familiar with.



Activities targeting specific rhetorical functions were generally well received by the PhD students, potentially because they noticed expressions that they had not used before. However, the observations of student work did not quite match the positive reception of these exercises. During the sessions it was clear that particularly exercises listing many expressions tended to foster rapid frequency comparisons at the expense of in-depth sentence analyses.

In view of the tendency to work very quickly and adopt an apparently surface-level approach in some exercises, it seems important to qualify claims like that by Jones and Haywood (2004: 272), who argue that: “The use of concordance lines as a way of studying lexis... is advantageous in that it requires a deep and thoughtful level of mental processing as students become involved in investigating for themselves the typical patterns of use of the target items.” Concordance lines certainly can have such an effect, but do not automatically generate deep level processing, for example when students are required to skim through them quickly. Developers of bundle exercises should thus beware of problems which may surface if activities highlight searches on a large number of bundles at the expense of appropriate levels of analysis.

The teaching and learning of bundles still remains relatively unexplored. There are several strands that need to be investigated in much greater detail. For instance, the fact that the students requested exercises on rhetorical functions suggests that this perspective is worth pursuing further. It would also be interesting to investigate the combination of bundles and rhetorical functions in a top-down and bottom-up approach to academic writing (Charles 2007). In addition, there is a need to explore the types of learning generated by different exercises in order find out more about what types of exercises may be appropriate at different levels. Finally, there is a need for studies which not only investigate the value of individual learning activities but also consider the sequencing of activities.

## References

- Anthony, L. 2010. *AntConc 3.2.1w*. <http://www.antlab.sci.waseda.ac.jp/software.html>, accessed 16/06/11.
- Aston, G. 2002. The learner as corpus designer. In B. Kettemann & G. Marko (eds), *Teaching and Learning by Doing Corpus Analysis*. Amsterdam: Rodopi, p. 9-25.
- Biber, D. 2006. *University Language. A corpus-based study of spoken and written English*. Amsterdam / Philadelphia: John Benjamins.
- Biggs, J.B. & K.F. Collins. 1982. *Evaluating the Quality of Learning: The SOLO taxonomy*. New York: Academic Press.
- Biggs, J. & C. Tang. 2007. *Teaching for Quality Learning at University*. 3<sup>rd</sup> edition. Maidenhead: McGraw-Hill Education/Society for Research into Higher Education & Open University Press.
- Byrd, P. & A. Coxhead. 2010. ‘On the other hand’: Lexical bundles in academic writing and in the teaching of EAP. *University of Sydney Papers in TESOL* 5: 31-64.
- Charles, M. 2007. Reconciling top-down and bottom-up approaches to graduate writing: Using a corpus to teach rhetorical functions. *Journal of English for Specific Purposes* 6: 289-302.
- Cortes, V. 2004. Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes* 23: 397-423.
- Cortes, V. 2006. Teaching lexical bundles in the disciplines: An example from a writing intensive history class. *Linguistics and Education* 17: 391-406.
- Coxhead, A. 2000. A new academic word list. *TESOL Quarterly* 34: 213-238.
- Coxhead, A. 2008. Phraseology and English for academic purposes: Challenges and opportunities. In S. Granger & F. Meunier (eds), *Phraseology in Foreign Language Learning and Teaching*. Amsterdam / Philadelphia: John Benjamins, p. 149-161.
- Ellis, N.C. 2008. Phraseology: The periphery and the heart of language. In S. Granger & F. Meunier (eds), *Phraseology in Foreign Language Learning and Teaching*. Amsterdam / Philadelphia: John Benjamins, p. 1-12.
- Flowerdew, L. 2004. The argument for using specialized corpora to understand academic and professional language. In U. Connor & T. Upton (eds), *Discourse in the Professions: Perspectives from corpus linguistics*. Amsterdam / Philadelphia: John Benjamins, p. 11-33.
- Flowerdew, L. 2008. Corpus linguistics for academic literacies mediated through discussion activities. In D. Belcher & A. Hirvela (eds), *The Oral-Literate Connection: Perspectives on L2 speaking, writing and other media interactions*. Ann Arbor: University of Michigan Press, p. 268-287.
- Flowerdew, L. 2009. Applying corpus linguistics to pedagogy: A critical evaluation. *International Journal of Corpus Linguistics* 14: 393-417.
- Flowerdew, L. 2010. Using corpora for writing instruction. In A. O’Keeffe & M. McCarthy (eds), *The Routledge Handbook of Corpus Linguistics*. London / New York: Routledge, p. 444-457.
- Granger, S. & F. Meunier (eds). 2008. *Phraseology in Foreign Language Learning and Teaching*. Amsterdam / Philadelphia: John Benjamins.
- Hyland, K. 2008a. ‘As can be seen’: Lexical bundles and disciplinary variation. *English for specific purposes* 27: 4-21.
- Hyland, K. 2008b. Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics* 18: 41-62.
- Jones, M. & S. Haywood. 2004. Facilitating the acquisition of formulaic sequences: An exploratory study in an EAP context. In N. Schmitt (ed.), *Formulaic Sequences*. Amsterdam / Philadelphia: John Benjamins, p. 269-292.
- Lee, D. & J. Swales. 2006. A corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora. *English for Specific Purposes* 25: 56-75.
- Lewis, M. 2000. Learning in the lexical approach. In M. Lewis (ed.), *Teaching Collocation: Further developments in the lexical approach*. Hove: Language Teaching Applications, p. 155-185.
- Mukherjee, J. & J-M. Rohrbach. 2006. Rethinking applied corpus linguistics from a language-pedagogical perspective: New departures in learner corpus research. In B. Kettemann & G. Marko (eds), *Planing, Gluing and Painting Corpora: Inside the applied corpus linguist’s workshop*. Frankfurt am Main: Peter Lang, p. 205-231.
- Nation, I.S.P. 2001. *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- Ngouchi, J. 2004. A genre analysis and mini-corpora approach to support professional writing by nonnative English speakers. *English Corpus Studies* 11: 101-110.
- O’Keeffe, A., M. McCarthy & R. Carter. 2007. *From Corpus to Classroom: Language use and language teaching*. Cambridge: Cambridge University Press.

- O'Sullivan, Í. 2007. Enhancing a process-oriented approach to literacy and language learning: The role of corpus consultation literacy. *ReCALL* 19: 269-286.
- Scott, M. & C. Tribble. 2006. *Textual Patterns: Key word and corpus analysis in language education*. Amsterdam / Philadelphia: John Benjamins.
- Seidlhofer, B. 2000. Operationalizing intertextuality: Using learner corpora for learning. In L. Burnard & T. McEnery (eds), *Rethinking Language Pedagogy from a Corpus Perspective*. Bern: Peter Lang, p. 207-223.
- Simpson-Vlach, R. & N.C. Ellis. 2010. An academic formulas list: New methods in phraseology research. *Applied Linguistics* 31: 487-512.
- Tribble, C. 2001. Small corpora and teaching writing. In M. Ghadessy, A. Henry & R.L. Roseberry (eds), *Small Corpus Studies and ELT: Theory and practice*. Amsterdam: John Benjamins, p. 381-408.
- Turnbull, J. & J. Burston. 1998. Towards independent concordance work for students: Lessons from a case study. *On-Call* 12(2): 10-21.
- Widdowson, H. G. 2000. On the limitations of linguistics applied. *Applied Linguistics* 21: 3-25
- Wray, A. 2002. *Formulaic Language and the Lexicon*. New York: Cambridge University Press.