CrossMark

Lang. Teach. (2015), **48.1**, 99–116 © Cambridge University Press 2013
doi:10.1017/S0261444813000037 First published online 26 February 2013

Corpus-based research and pedagogy in EAP: From lexis to genre

Lynne Flowerdew Hong Kong University of Science and Technology, Hong Kong SAR, China
lclynne@ust.hk

This plenary paper showcases current corpus-based research on written academic English, illustrating the tight links that exist between corpus research and pedagogic applications. I first explicate Sinclair's concept of the 'lexical approach', which underpins much corpus research and pedagogy. I then discuss studies which focus on individual lexical items, phraseology (including lexical bundles) and rhetorical functions. Finally, I examine studies that combine Swalesian move structure analysis with corpus analysis. Throughout the paper, I highlight the importance attached to lexis in both corpus research and pedagogic applications of corpus findings in the realm of written discourse in English for Academic Purposes (EAP), covering both general and more specific purposes.

1. Introduction

Corpora have played an increasingly important role in linguistic research and, to a lesser extent, teaching, over the past forty years. Some of this work involves grammar: witness Biber's (2006) notable work on grammatical variation between university registers and also the *Longman grammar of spoken and written English* (Biber et al. 1999), based on the analysis of a 40 million-word corpus spanning four different registers (conversation, newspaper, academic and fiction). Undoubtedly, while there are many excellent studies on grammatical categories yielding useful insights for applications to pedagogy, such as those just mentioned, many other initiatives concentrate on lexis. Why should this be so? One reason concerns the increasing attention paid to lexis since the 2000s. Nation's (2001) pioneering work on vocabulary served to redress the balance between the structuralist approach of the 1960s and current lexico-grammatical approaches, with the result that lexis is no longer regarded as the 'poor relation' of grammar. Indeed, some linguists, most notably Sinclair (1999), regard lexis as more important than grammar. Another reason for a greater tendency to focus on vocabulary rather than grammar in corpus analyses relates to the set-up of the corpus itself. For, as Halliday (2004) remarks, it is much easier to search for lexical items than for grammatical categories in a corpus, given that the main way to search a corpus is through software yielding

Revised version of a plenary address given at the 2012 International Conference on Applied Linguistics and Language Teaching (ALLT), National Taiwan University of Science and Technology, 20 April 2012.

https://doi.org/10.1017/S0261444813000037 Published online by Cambridge University Press

frequency lists of word types, i.e. the total number of different word forms. However, that is not to say that a corpus cannot be used for investigation of grammar. In cases where the focus is on grammar, the corpus has usually been tagged with grammatical categories to aid subsequent analysis. As a case in point, it would be quite laborious to search for instances of verbs in the past tense in an untagged corpus.

Moreover, as far as pedagogy is concerned, corpora have had far more impact on indirect applications, that is, on reference works such as dictionaries and grammars, than on direct ones (Stubbs 2004), where either concordance output is given to students or students search the corpus themselves through hands-on concordancing tasks. The reasons that concordancing has not 'percolated' down to classroom applications, or at least only at the level of individual institutions, include the difficulty of compiling corpora for a specific group of learners and the general lack of strategy training for students and teachers alike (Frankenberg-Garcia 2012).

In this presentation I concentrate on studies that focus primarily on lexis and that deal with academic writing for both English for General Academic Purposes (EGAP) and English for Specific Academic Purposes (ESAP). Of note is that, for the most part, they concern corpora which have been compiled at an institutional level by individuals (see Flowerdew 2004, 2010). I also focus on direct applications to pedagogy, illustrating the diverse ways in which corpora can be exploited for classroom use. After a brief explication of Sinclair's (1999) concept of the LEXICAL APPROACH, I look at some investigations and applications which take individual word types as their starting point. I then discuss several key studies which examine the phraseological behaviour of these individual word types, including those focusing on what are commonly known as LEXICAL BUNDLES. This is followed by analyses which consider lexis from a more functional perspective; for example, lexis such as *evidence* and *research* is often used in phrases which refer to the previous literature in a research paper. Finally, I look at research and pedagogic applications which take a Swalesian (1990, 2004) genre-based approach to corpus enquiries of academic discourse in which rhetorical functions are seen to operate at the level of MOVE STRUCTURE patterning. For instance, in Swales' oft-quoted CARS (CREATE-A-RESEARCH-SPACE) model for writing article introductions – Move 1: Establishing a territory; Move 2: Establishing a niche; Move 3: Occupying the niche – the rhetorical function of 'referring to the literature' would most commonly be found in Move 2, indicating a gap in the research field, identified through a phrase such as *little research has been carried out on*... In this exposition of how corpora have impacted research and pedagogy in the field of EAP, I therefore start with studies of a more bottom-up nature which consider individual word forms, moving on to studies taking a more top-down discourse-based perspective.

2. Concept of the lexical approach

In Sinclair's (1999) concept of the lexical approach, the lexical item is seen to have primacy over the grammar. And, indeed, it is this concept of a GRAMMATICALISED LEXIS which inspired the ELT lexical syllabus designed around frequency lists and their most important recurrent patterns (Willis 1990, 2003). As Willis (1990: vi) notes, for the first level of the COBUILD course the most frequent 700 words were selected from the corpus, these words accounting for around 70% of all English text. In addition, different senses of a word have

https://doi.org/10.1017/S0261444813000037 Published online by Cambridge University Press

different frequencies. For example, Sinclair & Renouf (1988), cited in J. Flowerdew (2009), note that the word *make* typically occurs in patterns such as *make decisions*, *make discoveries*, *make arrangements*. This delexical use of the verb *make* is more frequent than its concrete use, as in *make a cake*, and would therefore be given teaching priority. Sinclair & Renouf (1988) also point out that while the most frequent words typically have a range of uses, many of these uses are typically not covered in beginners' courses.

The underlying principle of the lexical syllabus is thus based on frequency. Nevertheless, this does not mean to say that grammar is neglected, as the lexis under investigation also incorporates the main grammatical forms. Sinclair (1991) provides corpus evidence to show that different senses of a word tend to occur in different lexico-grammatical environments, an observation also taken up in teaching materials. For example, Willis (1990: 38) notes the importance of *way*, the third most frequent noun in English, for language teaching. When *way* is used in the sense of 'direction', it is found in phrases such as 'A man asked me the way to St Paul's'. However, a more frequent sense is to indicate 'method' or 'means', typically encoded in the following lexico-grammar:

The best way to . . . is to . . .
One way of ___ing . . . is to . . . .

This interest in the lexical approach has also extended to EAP teaching. While corpora have proved valuable for identifying the salient lexis characteristic of different academic genres for use in pedagogy, there is some debate over whether there exists a core set of lexical items common to a variety of disciplines. This issue is taken up in the following section.

3. Lexical items

One key debate in the field concerns whether there exists a common core vocabulary across a range of academic texts and disciplines. Coxhead's (2000, 2011) corpus-based research of word families and vocabulary items in a 3.5-million-word corpus of academic texts from the disciplines of arts, commerce, law and science seems to support the common-core hypothesis, with the extraction of common vocabulary items forming an Academic Word List (AWL). Paquot's (2010) research also offers support for the AWL. However, Hyland & Tse's (2007) research points to a discipline-specific lexical repertoire. Other ESAP corpus-based studies, mostly using research articles (e.g. Chen & Ge 2007), in which wordlists from specialist disciplines are compared with Coxhead's academic wordlist, also favour sets of discipline-specific lexis (see Flowerdew 2012a for a summary of these). But as these studies rely on different types of corpora and use different software for analysis, it is not surprising that the findings differ and that this area is still fraught with controversies (Flowerdew 2012c). The situation is just as complex as far as phrases (e.g. *in order to*, *on the basis of*) are concerned. For example, the Academic Formulas List (AFL) developed by Simpson-Vlatch & Ellis (2010) and the Phrasal Expressions List drawn up by Martinez & Schmitt (2012) provide support for the AWL and show that high frequency phrases are not necessarily discipline-specific: findings which do not concur with those of the studies discussed in Section 5.1.

https://doi.org/10.1017/S0261444813000037 Published online by Cambridge University Press

so What if the gas were not ideal? A	**solution**	can still be found. Go back to
y the calculations required to obtain a	**solution.**	An average constant velocity
K-wall properties do not complicate the	**solution**	a great deal and give a better
Is dried and a water-based caustic soda	**solution**	Is added. Following a timed mix
he equation of state. A trial and error	**solution**	Is the simplest way to proceed.
the last trial. So this is an adequate	**solution:**	Notice how quickly the solution
or realm in. 4.1. A direct analytical	**solution**	is possible when the substance
ture is unknown, an explicit analytical	**solution**	is not possible. However, in the
to the desired answers. Attempting the	**solution**	with "shortcut" method is not
tion is obtained by adding a particular	**solution**	of the given equation to the
of copper. The electrolyte is an aqueous	**solution**	of and. Nothing happens until
11 not spontaneously be reduced in acid	**solution;**	Is positive. At the cathode,

Figure 1 Concordance data for **solution** (adapted from Mudraya 2006: 246)
Task: Study the concordance data and find the instances of the word **solution** used (1) in the general sense (e.g. ***solution of a problem***) and (2) in the (chemical) sense.

As far as pedagogy is concerned, Ward (2009) and Mudraya (2006), whose research studies focus on engineering textbooks, both advocate addressing vocabulary of a non-technical nature, mainly for the reason of high frequency and distribution across a range of texts. For example, Mudraya, finding, somewhat to her surprise, that the word *solution* occurs more frequently in its non-technical sense than its technical sense in a two-million-word corpus of engineering textbooks, advocates alerting students to both uses, as in the task shown in Figure 1.

While the focus of instruction seems to be on an individual word form, in actual fact the student would have to scrutinise the concordance lines syntagmatically (across) and paradigmatically (down) in order to work out the different senses of this lexical item. Such a task is rooted in Sinclair's (2004) model of language in which meaning is realised by the co-textual environment rather than residing inherently in individual words. Mudraya's task thus illustrates how the key specialised vocabulary in Coxhead's academic wordlist (Coxhead & Nation 2001) can be usefully extended to address the phraseological environment. Indeed, Coxhead (2008) also advocates expanding the AWL by identifying its common collocations and recurrent phrases. This phraseological aspect of language is the topic of the following section.

4. From lexis to phraseology and from phraseology to lexis

It goes without saying that the key language of written academic discourse is more than just individual vocabulary items. As Wible & Tsao (2011: 238) so eloquently put it: 'Tokens [the number of individual words] are valuable only as windows onto what they betoken. And basically what they betoken is patterns of word behaviours'. This leads us to a consideration of the phraseological behaviour of individual word forms, often conceptualised in terms

https://doi.org/10.1017/S0261444813000037 Published online by Cambridge University Press

of a PATTERN GRAMMAR (Hunston & Francis 2000), a view also espoused in Sinclair's lexical approach explicated in the previous section. Phraseology has been variously defined but, above all, can be seen as encompassing both 'collocation' and 'colligation'. While collocation is a well-known phenomenon, colligation has received less attention: it denotes the grammatical company that words keep, in terms of grammatical preferences rather than rule-based categories. However, it is only fairly recently that corpus studies have highlighted this important aspect of phraseology. For example, Byrd's (2007) examination of the lemma *require* in the AWL revealed *required* to be the most common word form, and, moreover, that this item displayed three possible complement patterns. Of note, here, is that *required* showed a distinct colligational preference for the infinitive, a useful finding for constructing EAP syllabi:

required + infinitive (most characteristic pattern)
required + 'that' clause
required + noun phrase complement

Meanwhile, Oakey's (2005: 178) research on the lemma *evaluate*, also occurring in the AWL, shows how phraseological patterning can differ depending on whether *evaluate* is found in the environment of active or passive voice. For example, in his corpus of 1,360 research articles from economics journals, two-thirds of the occurrences of *evaluate* were in the infinitive with *to*, most commonly occurring with various analysis tools which were the object of *use* when in the active, e.g. *This paper* ***uses*** *laboratory methods* ***to evaluate*** *whether price-fixing conspiracies break down*. . .. In contrast, when *evaluate* occurred with *use* in the passive, more lexical variation (*model/objective/functions/ration/assumption/measures/test/approach*) was found, e.g. . . .*Marshallian producer and consumer surplus measures* ***are used to evaluate*** *the welfare consequences of the given supply shift*.

The two examples above take lexis as the starting point for the phraseological enquiry. However, we can equally as well begin with a pattern and work back to the lexis. As Hunston & Francis put it (2000: 31), 'Sinclair's work is based on two possibly contradictory methodologies. One involves the researcher painstakingly investigating the phraseology of one lexical item after another. The other involves the use of a computer to list the most frequently-occurring word sequences'. Which of the two methodologies is used as the point of entry to the corpus depends heavily on the purpose and scope of the enquiry. I have pointed out (Flowerdew 2012a) that in a specialised corpus it may usually be more opportune to start with the lexis, one key reason being, as shown by Sinclair (2005), that the characteristic vocabulary of a special area features prominently in frequency lists. However, sometimes it may be more appropriate to start with a pattern, especially in instances where advanced students formulate their own linguistic queries.

In response to queries raised by their doctoral students, Lee & Swales (2006) devised a set of worksheets for hands-on concordancing. The worksheets had the following typical pattern, very much in the spirit of the inductive, hypothesis-testing approach usually associated with corpus-based pedagogy:

1. Guess and rank the frequencies, guess some exemplar word forms, and then check what the corpus says;

https://doi.org/10.1017/S0261444813000037 Published online by Cambridge University Press

2. Check some given sentences and correct where necessary;
3. For each pair of sentences, which alternative is better, and why?

One of their worksheet tasks was aimed at reinforcing the following useful pattern, based on students' request to investigate this.

ADJ/N/V for V-ing

(e.g. *appropriate for modelling. . .; devices for storing. . .; suited for studying. . .*):

Their workshop participants were also asked to rank the three word classes in terms of which was **most likely** to trigger the *for verb-ing* structure in academic writing (Answer: N, ADJ, V). Such tasks admirably illustrate how a corpus can be used to draw out the phraseological tendencies of academic writing by having students not only scrutinise the patterns but also pay attention to frequency counts. So here we have a pedagogic example of the grammar pattern as the point of entry with lexis as the end point.

Here I will digress for a moment to address the issue of frequency counts, a recurrent issue in course design. While frequency counts are important, they should not be seen as the be all and end all. As far back as the early 1990s, Widdowson (1991) queried a syllabus constructed entirely around frequency counts, revisiting this point several years later:

> . . . words and structures might be identified as 'pedagogically' core or nuclear and preferred as a prototype at a particular learning stage because of their coverage or their generative value, because they are catalysts which activate the learning process, whatever their status might be in respect to their actual occurrence in context of use. (Widdowson 2003: 87)

The same debate has been aired recently by Granger (2011: 135), who states that '. . . frequency needs to be counter-balanced by at least three other factors: learner variables, learnability and teachability'. Learner variables would include phenomena such as age, aptitude, motivation and learning style. The issue of learnability is concerned with psycholinguistic aspects such as memorisation, processing, storage and retrieval of phraseologies. Another factor to consider is teachability; for instance, it may well be more appropriate to teach the concrete rather than the metaphorical use of an item, even though the metaphorical use may occur more frequently.

There are a few EAP-oriented studies that go beyond simple frequency counts and also consider learnability and teachability. One such pedagogic initiative is that of Jones & Schmitt (2010) on corpus-derived materials for teaching discipline-specific vocabulary and phrases in academic seminars for international law and entrepreneurship. Of note is that their choice of key vocabulary, classified as technical, e.g. *regional body*; colloquial, e.g. *gut feeling*, or general, e.g. *take the initiative*, is not simply statistically derived but also takes account of input from both lecturers and students. Other key studies, one by Ellis, Simpson-Vlach & Maynard (2008) and a related study by Simpson-Vlach & Ellis (2010) on academic writing, move beyond the patterning of high-frequency lexical items to also take into account psycholinguistic salience and pedagogic relevance, as seen through teachers' eyes, to arrive at a pedagogic list of

https://doi.org/10.1017/S0261444813000037 Published online by Cambridge University Press

formulaic sequences, e.g. *in terms of*, *in order to*. Formulaic sequences, more commonly referred to as lexical bundles, are a well-researched area in corpus linguistics. The following section looks at how lexical bundles and the phraseology of lexical items have been categorised and taught from a functional perspective.

5. From lexical bundles and phraseology to functions

5.1 Lexical bundles

Phraseologies, one common type being LEXICAL BUNDLES, have also been classified functionally. Lexical bundles comprise a contiguous sequence of three or four words, identified purely through automatic means using a frequency-driven approach (cf. Biber, Conrad & Cortes 2004; Cortes 2004; Hyland 2008). Biber (2006: 134) notes that for a sequence of words to qualify as a lexical bundle, it would normally have to occur at least 40 times per million words in academic text. By way of example, Hyland's classification of lexical bundles contains three functional categories: research-oriented, text-oriented and participant-oriented, as outlined below (Hyland 2008: 13–14):

Research-oriented – help writers to structure their activities and experiences of the real world, e.g.:

Procedure (*the use of the*, *the operation of the*)
Quantification (*the magnitude of the*, *the surface of the*)

Text-oriented – concerned with organisation of the text and its meaning as a message or argument, e.g.:

Structuring signals – text-reflexive markers which organise stretches of discourse (*in the present study*, *in the next section*)
Framing signals – situate arguments by specifying limiting conditions (*in the case of*, *with respect to the*)

Participant-oriented – these are focused on the writer or reader of the text, e.g.:

Engagement features – address readers directly (*it should be noted that*, *it can be seen*)

Early initiatives on teaching lexical bundles are reported in Jones & Heywood (2004) and Cortes (2006), and are summarised in Eriksson (2012). Eriksson, in his own pedagogic application of corpora in teaching biochemistry doctoral students, used two small corpora of comparable size (c. 100,000 words), one consisting of 25 articles within the fields of biochemistry and biotechnology and the other made up of texts written by six doctoral students. His selection of bundles for teaching was, first, based on a comparison of bundles in these two corpora. Those found in the published texts but not in the Ph.D. texts were selected for pedagogic treatment. Of note is that Eriksson also consulted Hyland's (2008) list of lexical bundles (from biology and engineering texts) and Simpson-Vlach & Ellis' (2010) Academic Formulas List (AFL), a more general list. Eriksson notes that his cross-comparison showed

https://doi.org/10.1017/S0261444813000037 Published online by Cambridge University Press

that if Hyland's list and the AFL alone had been used, more than half of the bundles in his highly discipline-specific list would not have been found, thus underscoring the importance of specificity. In common with the tasks proposed by Lee & Swales (2006), Eriksson's exercises also adopted a hypothesis-testing, inductive approach (using the freely-available software *AntConc* (Anthony 2010)), as illustrated below (p. 207):

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.

What is quite surprising, though, is that in spite of the not inconsiderable research on lexical bundles, very few pedagogic applications are reported in the literature, the works mentioned above being notable exceptions. One reason for this lack of pedagogic uptake, as Byrd & Coxhead (2010) have pointed out, could well be the lack of information on the use of bundles in context. They investigated *on the basis of*, a key bundle found in many lists in the literature. Their finer analysis of this bundle in the AWL reveals it to have three possible uses:

a. Used at beginning of a sentence to provide a transition and specify methods used to carry out a process, e.g. ***On the basis of*** *his analyses, Clyne also identifies a number of. . .*
b. Used as adverbial of reason in a passive sentence or clause to explain the way that a decision was made or data handled, e.g. *Meanwhile, unskilled and unassisted migrants, most notably from Southern Europe were accepted* ***on the basis of*** *nomination by relatives in Australia. . ..*
c. Used with an adverbial for strengthening or diminishing the force of an utterance, e.g. *Only for L. notosaurus was the decision on its specific distinction made* ***solely on the basis of*** *allotropic data.* (adapted from Byrd & Coxhead 2010: 53–54)

5.2 Phraseology

More often than not, corpus-driven learning centred on key rhetorical functions of academic writing takes lexis as the starting point, extending this to phraseology. One early key initiative

https://doi.org/10.1017/S0261444813000037 Published online by Cambridge University Press

following this principle is that of Thurstun & Candlin (1998a, b) on writing in English for General Academic Purposes (EGAP). In their corpus materials based on concordance output, within each broad rhetorical function (e.g. stating the topic, referring to the literature, reporting the research of others), salient words (e.g. *issue*, *claim*, *identify*) are examined following the four-step activity below. As we can see, these encourage students to focus first and foremost on the meaning and pattern:

> **Step 1**: LOOK at concordances for the key term and words surrounding it, thinking of meaning.
> **Step 2**: FAMILIARISE yourself with the patterns of language surrounding the key term by referring to the concordances as you complete the tasks.
> **Step 3**: PRACTISE key terms without referring to the concordances.
> **Step 4**: CREATE your own piece of writing using the terms studied to fulfil a particular function of academic writing. (adapted from Thurstun & Candlin 1998b: 272)

In Step 1 students are asked to examine the concordance output of a key word and to reflect on how it is used in context. For example, the item *issue* is often used in stating the topic of the writing, e.g. *The issue for consideration is the selective treatment of disabled new-born babies*. The tasks in Step 2 require more detailed examination of the concordance output for *issue*, moving from lexis to pattern. Students are asked to group the adjectives often found before *issue*. Such a task not only targets collocation, but also alerts them to the concept of semantic association (also known as semantic preference), i.e. *issue* tends to be modified by words belonging to a restricted range of semantic sets such as importance or quantity.

> Step 2: **FAMILIARISE** (key word ***issue***)
>
> 1. Which *adjectives* are often used before **issue**?
> Group 1 (subject area) ***economic*** issue
> Group 2 (type/importance) ***key*** issue
> Group 3 (quantity) ***single*** issue
> 2. Which prepositions commonly follow **issue**?
> (adapted from Thurstun & Candlin 1998a: 1–3)

Functionally-oriented material for EGAP has also been produced by Charles (2007, 2011) and Bloch (2008, 2009, 2010). Bloch's tasks, targeted at undergraduate students, are based on his own in-depth corpus research on the use of reporting verbs in academic writing. However, unlike the pen-and-paper tasks proposed by Thurstun & Candlin discussed above, Bloch's corpus-based tasks are accessed through his own specially designed user-friendly program. The interface allows students to choose between searching for a specific word, which will limit the concordance display to a manageable 20 samples, or to search by concept, which leads the student through five prompt categories, as outlined below for the rhetorical purpose of 'making a claim'.

> Think about the claim you want to make and then answer the following questions:
>
> Do you want an **Integral** or **Non-integral** example?
> Do you want an **Informative** or a **Descriptive** example?
> Do you want an example from the point of view of the author you are citing or yourself as the writer?

https://doi.org/10.1017/S0261444813000037 Published online by Cambridge University Press

You can express an attitude towards the truth of a claim:
(1) **Positive** (2) **Negative** (3) **Unclear**
You can modify or hedge how strong that attitude is:
(1) **Strong** (2) **Moderate** (3) **Weak**

(adapted from Bloch 2008: 199)

If we compare Bloch's exercise typology for 'making a claim' with that of Eriksson (2012), discussed in Section 5.1, for 'qualifying claims' from a lexical bundle perspective, we can see the diverse activities to which corpus data can lend themselves. However, how a corpus is used is highly dependent on the level and language proficiency of the students. Eriksson's doctoral students are advanced, and consulting a corpus from their own discipline, biochemistry; the use of hands-on concordancing activities with minimum guidance from the teacher would be appropriate for this type of student. Meanwhile, Bloch, to cater to his lower-level undergraduate students, has mediated the raw corpus data of academic writing in a number of ways, e.g. limiting the concordance output to 20 samples and, most importantly, allowing students to search by rhetorical function, which is particularly useful as it supplies the student with the key lexis. One criticism levelled against corpus work with lower-level students is that they often don't know which key item to search, so this type of pedagogic mediation overcomes this obstacle.

Another corpus researcher/practitioner who, like Eriksson, is concerned with hands-on corpus activities targeted at doctoral students, is Charles (2007, 2011). In common with Bloch and Eriksson, Charles' activities are also grounded in her corpus analysis, in this case a 500,000-word corpus of theses, thus underscoring the close link between corpus research and pedagogy.

In Charles (2007: 296) details are given of hands-on tasks designed to familiarise students with the general function of *defending your work against criticism*. First, students search on ***while*** in sentence-initial position with the context terms ***appear* / seem* may*** occurring within an eight-word span, within the same sentence. Such a search reveals that ***while*** at the beginning of the sentence is used to construct a concession where the importance of the conceded proposition is downgraded and emphasis put on the writer's defence. This two-part pattern, i.e. anticipated criticism → defence, and its realisation using signals of apparent concession, contrast and justification, is illustrated in Table 1.

Another feature of Charles' materials is that she approaches this two-part function by first using a top-down approach (see Durrant & Matthews-Aydinli 2011, who have argued for a 'function-first' approach in analysing formulaic language, i.e. phraseology). Students are provided with a suite of pen-and-paper worksheet activities to sensitise them to the extended discourse properties of this two-part function. She then supplements these activities with a more bottom-up approach by having students search the corpus to identify typical lexico-grammatical patterns realising these functions. Other corpus linguists have looked at functions within the overall move structure patterning of text, very often taking the lexical item as the starting point (see the Introduction for a brief explanation), which is the focus of the following section.

https://doi.org/10.1017/S0261444813000037 Published online by Cambridge University Press

Table 1 Anticipated criticism and writer's defence (adapted from Charles 2007: 294)

Extract	Anticipated criticism	Writer's defence
	While I acknowledge that in some cases the distinction between institutions and groups **may seem** rather arbitrary	Such political actor subjects are not the focus of attention in this thesis.

6. From functions to move structures

The vast majority of research and the pedagogic applications carried out on move structure patterning focuses on research articles and legal genres such as the formal essay. A key reason for this is that only those genres which display fairly prototypical move structure patterning, as is the case for the CARS model (see Section 1), would lend themselves to such treatment. The second reason concerns accessibility of texts (cf. Flowerdew 2004). It would be far easier to compile a corpus of research articles than Ph.D. theses, for instance, as the latter have certain copyright restrictions and belong to a 'semi-occluded' genre (Swales 2004). One genre-inspired study on Ph.D. thesis writing is that by J. Flowerdew & Forest (2009). They applied Swales' (1990: 141) CARS (create-a-research-space) model, originally proposed for introductions in academic research articles, to Ph.D. literature reviews in the field of applied linguistics. It is worth noting that they start from a bottom-up perspective with their investigation of the patterning of the keyword *research* and then ask to which particular move structures the different phraseologies of *research* correspond. For example, it was found for the GAP INDICATING move that *research* was involved in two common patterns: *There has been little research/little research has been done*; *Further research is needed/called for* (p. 27).

Kanoksilapathan (2005, 2007) investigated the lexico-grammatical characteristics of moves in a corpus of 60 research articles from biochemistry. Kanoksilapathan first identified the main moves in each main section of the research article. For example, the following four types of main moves were identified in the Results section:

Move 8: Restating methodological issues
Move 9: Justifying methodological issues
Move 10: Announcing results
Move 11: Commenting results

Each move was then broken down into steps, with the 'Commenting results' move ascribed the following five steps:

Move 11: Commenting results
  Step 1: Explaining results
  Step 2: Generalising/interpreting results
  Step 3: Evaluating results
  Step 4: Stating limitations
  Step 5: Summarising

(adapted from Kanoksilapathan 2007: 76–81)

https://doi.org/10.1017/S0261444813000037 Published online by Cambridge University Press

However, the second stage of Kanoksilapathan's research differs quite markedly from that of J. Flowerdew & Forest as she then takes a multimethod approach, combining move structure analysis with a modified version of Biber's (1988) multidimensional analysis for identifying functionally-motivated dimensions of text; for example, what Kanoksilapathan terms 'evaluative stance' would correspond to Biber's Dimension 4, 'Overt expression of argumentation'. In Biber's system the dimensions are identified in terms of grammatical items, e.g. 'that' complement clauses or the extraposed 'it' construction, but Kanoksilapathan looks at these features in their immediate co-textual environment, which reveals the lexis whereby authors express their comments or attitudes, e.g. '***It*** *is interesting that* the experiments in this paper were all carried out using assays for genetic interference in. . .' (p. 94). In sum, Kanoksilapathan's research thus moves from a top-down to a more bottom-up approach in which lexis does have a role to play, but one which is more akin to that in Halliday's system in which lexis is conceived 'as the most delicate grammar'.

As for pedagogic applications which make use of corpus-based genre materials, Bianchi & Pazzaglia (2007) propose a suite of activities for helping students to write psychology research articles in English. First, students were asked to subdivide a written article chosen by themselves into moves and then annotate it using a functional and meta-communicative coding system devised by the authors. This was followed by examination of concordances of keywords related to the concept of research, e.g. *study/studies*, *experiment/experiments*, *literature* and *research/researches*. Students also examined the lexico-grammatical environment of these keywords, noting verb and tense usage, e.g. *The aim of this research was to analyse*; *In this research we analysed* (p. 271). This corpus application can thus be seen to mirror the methodology adopted in J. Flowerdew & Forest's study outlined above, with its bottom-up approach and focus on key lexis.

Another study with pedagogic applications is that by Chang & Kuo (2011), who based their EAP course, designed specifically to help postgraduate students write research articles, on a corpus of 60 research articles from computer science. In fact, their proposed course, like Bloch's described in the previous subsection, relies on a specially-constructed website, in this case using *Moodle*, a public-domain tool for presenting instructional materials on the web. Their online EAP platform, organised around the major sections of the research article (Introduction, Materials and Methods, Results and Discussions, Conclusions and Abstract), consists of information structure units and language use units. The information structure units present the rhetorical moves and common move patterns based on the authors' previous genre analysis. The language use unit is supported by online writing aids such as dictionaries and also provides links to a collocation builder and concordancer.

The pedagogic applications by Bianchi & Pazzaglia (2007) and Chang & Kuo (2011) described above both use corpora of research articles for training postgraduate students to write research articles. Such corpora have also been exploited to help doctoral students with thesis writing (Thompson & Tribble 2001; Charles 2011). While it must be acknowledged that there are some differences in rhetorical structure between these two related genres, judicious use of research articles can also be of value in teaching thesis writing. In workshops on writing the Discussion section of theses, I have used the freely-available Hong Kong PolyU Corpus of Research Articles (CRA) (see Lin & Evans 2012) consisting of about five million words covering 39 disciplines. Of note is the fact that the CRA uses the search engine *ConcGram*,

https://doi.org/10.1017/S0261444813000037 Published online by Cambridge University Press

Table 2 Verbs occurring in genre moves in law cases (Bhatia et al. 2004: 214)

	Frequency			
Genre Move	Submit	Dismiss	Reject	Grant
1 Facts/Stating history of the case	75	**47**	12	**82**
2 Presenting argument	**263**	6	9	51
3 Deriving ratio decidendi	5	16	**44**	**80**
4 Pronouncing judgment	3	**42**	9	16
Total	**346**	**111**	**74**	**229**

which allows for both positional and constituency variation (Cheng, Greaves & Warren 2006; Greaves 2009).

To take an example, I devised the following hypothesis-testing task to familiarise students with lexical phrases for commenting on results. Starting with the lexical item 'surprising', I asked students to write a sentence incorporating this word. The following is an example of the kind of writing students produced: 'It is surprising (to find) that. . .' However, inspection of the corpus data, sorted 'left', revealed that the preferred patterning for this word was with negative polarity e.g. *This phenomenon is not surprising as such, as they also appear in polymer solutions. . .; This is not surprising as the CANE model. . .*, thus echoing Halliday's (1991) corpus evidence for the unequal distribution of polarity across different verbs, although in this case an adjective is involved. Moreover, the concordance lines for 'surprising', right-sorted, revealed the co-occurrence of this item with the expression 'given that'. This kind of empirical data alerts students to the fact that language has certain phraseological tendencies, depending on the genre under investigation.

Meanwhile, Bhatia, Langton & Lung (2004) have shown how lexical verb choice operates at the level of move structure in the highly conventionalised genre of law cases, as shown in Table 2. We can see that seemingly synonymous verbs such as *dismiss* and *reject* have a clear preference for different move structures, and that it would be necessary, as Bhatia et al. (2004) point out, to consult law experts on their pragmatic difference as corpora cannot (usually) provide this kind of information.

Law cases are highly relevant for the Problem-Question genre of writing in the academy, a means of acculturating law students into professional practices. In the Problem-Question genre, law students are expected to identify a range of issues and relevant law from the facts, create and develop arguments to draw conclusions on likely outcomes, and demonstrate deductive reasoning, the same skills required for the analysis of law cases. As noted by Bhatia et al. (2004), deductive reasoning, which plays a major role in this genre, requires mastery of various hedging devices which are especially important in the application to facts section. To sensitise law students to this, Bhatia et al. suggest having students identify and classify hedges by function, form or grammar according to move structure, using worksheets based on concordance output extracted from law cases.

https://doi.org/10.1017/S0261444813000037 Published online by Cambridge University Press

received Brian's letter.	**Assuming**	the offer does remain open, Brian's Thursda
proceeding on the latter	**assumption**	In order to discuss the law related to
to discuss. I now have to	**consider**	whether B's message, left on the answerphone
third party. Bata v. Bata.	**Considering**	first the story about The BCDs, can the ba
With	**Regard to**	contracts *ex facie* illegal it is necessary to
second part of the story.	**Regarding**	the potential claim of Evangeline, it is subm
consented to it. The key	**issue**	here is what caused the injury; was it Geoffrey's
At	**issue**	here is whether Neil will have a cause of action

Figure 2 Signals in an opening-type move (adapted from Weber 2001: 17)

The formal essay is another common legal genre in the academy (see Weber 2001). As in the initial rhetorical awareness-raising activities adopted by both Charles and Bianchi & Pazzaglia, as described previously, Weber's students were first inducted into the genre of legal essays by reading through whole essays taken from the University of London LL.B. examinations written by native speakers, and identifying some of the prototypical rhetorical features, e.g. identifying and/or delimiting the legal principle involved in the case. Weber's students were then asked to identify any lexical expressions which seemed to correlate with these features. The students then consulted the corpus of the legal essays to verify the patterns they had noted, in other words, to test their hypotheses, and also to pinpoint any regularities in lexico-grammatical expressions. Items such as *assume*, *consider*, *regard* and *issue*, in various constructions, were all found to act as signals in an opening-type move, delimiting the case under consideration before the principle involved in it was defined, as exemplified in Figure 2.

All the above studies combining genre move structure analysis and corpus analysis bear witness to the fact that corpora are no longer just consulted for lexico-grammatical patterns at the sentence level (Flowerdew 2005), but are now being fruitfully used to investigate and teach more discourse-level phenomena (see Handford 2010; Flowerdew 2012b).

7. Conclusion

In sum, this plenary paper has presented an overview of current corpus-based research on written academic English, encompassing individual vocabulary items, phraseology, rhetorical functions and Swalesian move structure genre analysis, together with a discussion of how the findings have impacted pedagogy. One observation that can be made is that there is a tight bond between research and pedagogy in corpus work, with one informing the other. As I emphasised at the beginning of this talk, lexis assumes a major role in corpus endeavours, usually functioning as the entry point into corpus investigations, but this is not always the case. I have also commented on bottom-up and more top-down approaches to corpus analysis, with many studies adopting both. While most of the corpus studies on EAP reported in this plenary paper tend to revolve around small-scale projects carried out at an individual level, this situation is now changing. Much larger-scale projects on student writing in the academy across a wide range of genres, disciplines and levels have very recently been completed; we have seen the compilation of MICUSP, the Michigan Corpus of Upper-level Student

https://doi.org/10.1017/S0261444813000037 Published online by Cambridge University Press

Papers (Römer & Brook O'Donnell 2011) and BAWE, British Academic Written English Corpus (Nesi 2011; Nesi & Gardner 2012). Another noteworthy large-scale project is that by Cotos et al. (2012) examining the issue of interdisciplinarity in a corpus of 900 research articles across 30 disciplines. Research findings from these corpora will undoubtedly provide insightful and important contributions to the field of EAP for both research and pedagogy.

Acknowledgments

I wish to thank the three anonymous referees for their very careful reading of the article and their valuable suggestions for revisions.

References

Anthony, L. (2010). *AntConc* 3.2.1W. www.antlab.sci.waseda.ac.jp/software.html
Bhatia, V. K., N. Langton & J. Lung (2004). Legal discourse: Opportunities and threats for corpus linguistics. In U. Connor & T. Upton (eds.), 203–231.
Bianchi, F. & R. Pazzaglia (2007). Student writing of research articles in a foreign language: Metacognition and corpora. In R. Facchinetti (ed.), *Corpus linguistics 25 years on*. Amsterdam: Rodopi, 259–287.
Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.
Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. Amsterdam: John Benjamins.
Biber, D., S. Conrad & V. Cortes (2004). *If you look at. . ..*: Lexical bundles in university teaching and textbooks. *Applied Linguistics* 25.3, 371–405.
Biber, D., S. Johansson, G. Leech, S. Conrad & E. Finegan (1999). *The Longman grammar of spoken and written English*. London: Longman.
Bloch, J. (2008). *Technologies in the second language composition class*. Ann Arbor, MI: University of Michigan Press.
Bloch, J. (2009). The design of an online concordancing program for teaching about reporting verbs. *Language Learning & Technology* 13.1, 59–78.
Bloch, J. (2010). A concordance-based study of the use of reporting verbs as rhetorical devices in academic papers. *Journal of Writing Research* 2.2, 219–244.
Byrd, P. (2007). Collocations and recurrent phrases in the Academic Word List. Paper presented at 41st Annual TESOL Convention, Seattle, Washington, 23 March.
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.
Chang, C.-F. & C.-H. Kuo (2011). A corpus-based approach to online materials development for writing research articles. *English for Specific Purposes* 30, 222–234.
Charles, M. (2007). Reconciling top-down and bottom-up approaches to graduate writing: Using a corpus to teach rhetorical functions. *Journal of English for Academic Purposes* 6.4, 289–302.
Charles, M. (2011). Using hands-on concordancing to teach rhetorical functions: Evaluation and implications for EAP writing classes. In A. Frankenberg-Garcia, L. Flowerdew & G. Aston (eds.), 26–43.
Chen, Q. & C. Ge (2007). A corpus-based lexical study on frequency and distribution of Coxhead's AWL word families in medical research articles. *English for Specific Purposes* 26.4, 502–514.
Cheng, W., C. Greaves & M. Warren (2006). From n-gram to skipgram to concgram. *International Journal of Corpus Linguistics* 11.4, 411–433.
Connor, U. & T. Upton (eds.) (2005). *Discourse in the professions: Perspectives from corpus linguistics*. Amsterdam: John Benjamins.
Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes* 23, 397–423.

https://doi.org/10.1017/S0261444813000037 Published online by Cambridge University Press

Cortes, V. (2006). Teaching lexical bundles in the disciplines: An example from a writing intensive history class. *Linguistics and Education* 17, 391–406.
Cotos, E., S. Huffman, S. Link, T. Paben & A. Saricaoglu (2012). Cross-disciplinary move schema for research article methods sections. Paper presented at AAAL Conference, Boston, MA, 27 March 2012.
Coxhead, A. (2000). A new academic wordlist. *TESOL Quarterly* 34.2, 213–238.
Coxhead, A. (2008). Phraseology and English for academic purposes. Challenges and opportunities. In F. Meunier & S. Granger (eds.), 149–161.
Coxhead, A. (2011). The academic word list 10 years on: Research and teaching implications. *TESOL Quarterly* 45.2, 355–362.
Coxhead, A. & P. Nation (2001). The specialised vocabulary of English for academic purposes. In J. Flowerdew & M. Peacock (eds.), *Research perspectives on English for academic purposes*. Cambridge: Cambridge University Press, 252–267.
Durrant, P. & J. Matthews-Aydinli (2011). A function-first approach to identifying formulaic language in academic writing. *English for Specific Purposes* 30.1, 58–72.
Ellis, N., R. Simpson-Vlach & C. Maynard (2008). Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics and TESOL. *TESOL Quarterly* 42.3, 375–396.
Eriksson, A. (2012). Pedagogical perspectives on bundles: Teaching bundles to doctoral students of biochemistry. In J. Thomas & A. Boulton (eds.), *Input, process and product: Developments in teaching and language corpora*. Brno, Czech Rep.: Masaryk University Press, 195–211.
Flowerdew, J. (2009). Corpora in language teaching. In M. Long & C. Doughty (eds.), *The handbook of language teaching*. London: Wiley-Blackwell, 327–350.
Flowerdew, J. & R. Forest (2009). Schematic structure and lexico-grammatical realization in corpus-based genre analysis. In M. Charles, D. Pecorari & S. Hunston (eds.), *Academic writing: At the interface of corpus and discourse*. London: Continuum, 15–36.
Flowerdew, L. (2004). The argument for using specialised corpora to understand academic and professional language. In U. Connor & T. Upton (eds.), 11–33.
Flowerdew, L. (2005). An integration of corpus-based and genre-based approaches to text analysis in EAP/ESP: Countering criticisms against corpus-based methodologies. *English for Specific Purposes* 24, 321–332.
Flowerdew, L. (2010). Using corpora for writing instruction. In A. O'Keeffe & M. McCarthy (eds.), 444–457.
Flowerdew, L. (2012a). *Corpora and language education*. London: Palgrave Macmillan.
Flowerdew, L. (2012b). Corpus-based discourse analysis. In J. P. Gee & M. Handford (eds.), *The Routledge handbook of discourse analysis*. London: Routledge, 174–187.
Flowerdew, L. (2012c). English for academic purposes. In C. Chapelle (ed.) *The encyclopedia of applied linguistics*. Oxford, UK: Wiley-Blackwell.
Frankenberg-Garcia, A. (2012). Raising teachers' awareness of corpora. *Language Teaching* 45.4, 475–489.
Frankenberg-Garcia, A., L. Flowerdew & G. Aston (eds.) (2011). *New trends in corpora and language learning*. London: Continuum.
Granger, S. (2011). From phraseology to pedagogy: Challenges and prospects. In T. Herbst, S. Faulhaber & P. Uhrig (eds.), *The phraseological view of language: A tribute to John Sinclair*. Berlin: De Gruyter Mouton, 123–146.
Greaves, C. (2009). *ConcGram 1.0: A phraseological search engine*. Amsterdam: John Benjamins.
Halliday, M. A. K. (1991). Corpus studies and probabilistic grammars. In K. Aijmer & B. Altenberg (eds.), *Advances in corpus linguistics*. Amsterdam: Rodopi, 30–43.
Halliday, M. A. K. (2004). The spoken language corpus: A foundation for grammatical theory. In K. Aijmer & B. Altenberg (eds.), *Advances in corpus linguistics*. Amsterdam: Rodopi, 11–38.
Handford, M. (2010). What can a corpus tell us about specialist genres? In A. O'Keeffe & M. McCarthy (eds.), *The Routledge handbook of corpus linguistics*. London: Routledge, 255–269.
Hong Kong PolyU Corpus of Research Articles (CRA). http://rcpce.engl.polyu.edu.hk/RACorpus/default.htm
Hunston, S. & G. Francis (2000). *Pattern grammar*. Amsterdam: John Benjamins.
Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes* 27.1, 4–21.

https://doi.org/10.1017/S0261444813000037 Published online by Cambridge University Press

Hyland, K. & P. Tse (2007). Is there an 'academic vocabulary'? *TESOL Quarterly* 41, 235–253.
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: John Benjamins, 269–292.
Jones, M. & N. Schmitt (2010). Developing materials for discipline-specific vocabulary and phrases in academic seminars. In N. Harwood (ed.), *English language teaching materials: Theory and practice*. Cambridge: Cambridge University Press, 225–250.
Kanoksilapathan, B. (2005). Rhetorical structure of biochemistry research articles. *English for Specific Purposes* 24.3, 269–292.
Kanoksilapathan, B. (2007). Rhetorical moves in biochemistry research articles. In D. Biber, U. Connor & T. Upton (eds.), *Discourse on the move*. Amsterdam: John Benjamins, 73–119.
Lee, D. & J. Swales (2006). A corpus-based EAP course for NNS doctoral students: Moving from available specialised corpora to self-compiled corpora. *English for Specific Purposes* 25, 56–75.
Lin, L. & S. Evans (2012). Structural patterns in empirical research articles. *English for Specific Purposes* 31, 150–160.
Martinez, R. & N. Schmitt (2012). A phrasal expression list. *Applied Linguistics* 33.3, 299–320.
Meunier, F., S. De Cock, G. Gilquin & M. Paquot (eds.) (2011). *A taste for corpora*. Amsterdam: John Benjamins.
Meunier, F. & S. Granger (eds.) (2008). *Phraseology in foreign language learning and teaching*. Amsterdam: John Benjamins.
Mudraya, O. (2006). Engineering English: A lexical frequency instructional model. *English for Specific Purposes* 25.2, 235–256.
Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
Nesi, H. (2011). BAWE: An introduction to a new resource. In A. Frankenberg-Garcia, L. Flowerdew & G. Aston (eds.), 213–228.
Nesi, H. & S. Gardner (2012). *Genres across the disciplines. Student writing in higher education*. Cambridge: Cambridge University Press.
Oakey, D. (2005). Academic vocabulary in academic discourse: The phraseological behaviour of EVALUATION in economics research articles. In E. Tognini Bonelli & G. Del Lungo Camiciotti (eds.), *Strategies in academic discourse*. Amsterdam: John Benjamins, 169–183.
Paquot, M. (2010). *Academic vocabulary in learner writing*. London: Continuum.
Römer, U. & M. Brook O'Donnell (2011). From student hard drive to web corpus (part 1): The design, compilation and genre classification of the Michigan Corpus of Upper Student Papers (MICUSP). *Corpora* 6.2, 159–177.
Simpson-Vlach, R. & N. Ellis (2010). An academic formulas list: New methods in phraseological research. *Applied Linguistics* 31.4, 487–512.
Sinclair, J. McH. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
Sinclair, J. McH. (1999). The lexical item. In E. Weigand (ed.), *Contrastive lexical semantics*, Amsterdam: John Benjamins, 1–24.
Sinclair, J. McH. (2004). New evidence, new priorities, new attitudes. In J. Sinclair (ed.), *How to use corpora in language teaching*. Amsterdam: John Benjamins, 271–299.
Sinclair, J. McH. (2005). Corpus and text: Basic principles. In M. Wynne (ed.), *Developing linguistic corpora: A guide to good practice*. Oxford Text Archive. Available at http://ahds.ac.uk/linguistic-corpora/, 1–21.
Sinclair, J. McH. & A. Renouf (1988). A lexical syllabus for language learning. In R. Carter & M. McCarthy (eds.), *Vocabulary and language teaching*. London: Longman, 140–160.
Stubbs, M. (2004). Language corpora. In A. Davies & C. Elder (eds.), *The handbook of applied linguistics*. Malden, MA: Blackwell, 106–132.
Swales, J. (1990). *Genre analysis*. Cambridge: Cambridge University Press.
Swales, J. (2004). *Research genres*. Cambridge: Cambridge University Press.
Thompson, P. & C. Tribble (2001). Looking at citations: Using corpora in English for academic purposes. *Language Learning & Technology* 5.3, 91–105.
Thurstun, J. & C. Candlin (1998a). *Exploring academic English*. Macquarie University: NCELTR.
Thurstun, J. & C. Candlin (1998b). Concordancing and the teaching of vocabulary of academic English. *English for Specific Purposes* 17.3, 267–280.
Ward, J. (2009). A basic engineering word list for less proficient foundation engineering undergraduates. *English for Specific Purposes* 28, 170–182.

https://doi.org/10.1017/S0261444813000037 Published online by Cambridge University Press

Weber, J.-J. (2001). A concordance- and genre-informed approach to ESP essay writing. *ELT Journal* 55.1, 14–20.
Wible, D. & N.-L. Tsao (2011). Towards a new generation of corpus-derived lexical resources for language learning. In F. Meunier et al. (eds.), 237–255.
Widdowson, H. G. (1991). The description and prescription of language. *Linguistics and language pedagogy: The state of the art*. Washington, DC: Georgetown University Press, 11–24.
Widdowson, H. G. (2003). *Defining issues in English language teaching*. Oxford: Oxford University Press.
Willis, D. (1990). *The lexical syllabus: A new approach to language teaching*. London: HarperCollins.
Willis, D. (2003). *Rules, patterns and words: Grammar and lexis in English language teaching*. Cambridge: Cambridge University Press.

Lynne Flowerdew works in the Center for Language Education at the Hong Kong University of Science and Technology, where she coordinates courses and workshops for postgraduate research students. Her main research interests include corpus linguistics, discourse analysis, genre analysis and ESP. She has published widely in these areas; her most recent book is *Corpora and language education* (Palgrave Macmillan 2012).

https://doi.org/10.1017/S0261444813000037 Published online by Cambridge University Press