

# **An analysis of the usefulness of formulaic sequences and related acquisition exercises in an EFL academic writing textbook**

Lewis Murray

The high incidence of formulaic sequences (FS), defined as “combinations of at least two words favoured by native speakers (NS) in preference to an alternative combination which could have been equivalent had there been no conventionalization” (Erman & Warren, 2000), is well recognised within discourse (Boers & Lindstromberg, 2012; Meunier, 2012; Wray, 2002). It has received particular attention within written academic discourse, where it is seen to be especially prevalent (Adel & Erman, 2012; AlHassan & Wood, 2015; Ellis et al, 2008), and has been the focus of numerous studies seeking to establish which FS occur most frequently and might therefore be of greatest value to second language (L2) learners (Hsu, 2014; Hyland, 2008a; Simpson-Vlach & Ellis, 2010). Studies examining the usefulness of FS within EFL textbooks have tended not to conclude favourably, instead raising questions regarding the basis of the authors’ lexical selection (Koprowski, 2005; McAleese, 2013; Wood, 2010). Research also indicates that L2 learners’ successful acquisition of FS may be enhanced through explicit attention, in terms of noticing and utilization (AlHassan & Wood, 2015; Peters & Pauwels, 2014), but here again studies suggest that textbooks may be failing learners (Woods, 2010). In light of this, the current study reports on a textbook analysis of ‘Writing Essays: from paragraph to essay’ (Zemach & Ghulldu, 2011), to examine the extent to which such claims, regarding selection and utilization of FS, apply therein.

## Literature

Since Sinclair's early corpus studies suggested the phraseological, formulaic nature of language (Sinclair, 1991), numerous subsequent studies have further revealed this to be the case (see, for example, Conklin & Schmitt, 2008; Wray, 2002). Rather than make use of the generative facility available, allowing for numerous ways of conveying a single meaning, corpus studies have shown that within idiomatic language use, the incidence of repeated 'chunks' is significant (Greaves & Warren, 2010). Research also suggests that such language is genre-sensitive; certain sequences occur with significantly greater frequency within particular discourse genres (Greaves & Warren, 2010; Schmitt, 2010). It is thought, therefore, that proficient use of idiomatic language may contribute to one's identification as a member of a particular discourse community (Millar, 2011; Wray, 2000). Additionally, for those well-versed in the discourse of a particular community, appropriate use of such language is thought to afford processing advantages (Schmitt, 2010; Wray, 2009). As such, it is perhaps no surprise that it has also received much attention within the field of second language acquisition (SLA). However, research therein suggests that L2 learners' facility with and recognition of FS may be problematic, with studies indicating, for example, that acquisition may not be incidental, but rather that FS need to be taught explicitly (Adel & Erman, 2012; Peters & Pauwels, 2014).

With regard to the teaching of FS, a number of issues exist, among which is the question of which to teach. With FS estimated to account for between a third and a half of any given discourse type (Conklin & Schmitt, 2008), selecting those most beneficial for L2 learners must surely be a priority. As noted above, FS have been found to be genre-specific, and this has led a number of researchers to focus their efforts on uncovering the most useful items within a given genre (Martinez & Schmitt, 2012). One field that has received particular attention is academic writing, where FS are found to be especially prevalent (Hyland, 2008b) resulting in the

publication of numerous lists of those deemed particularly useful (Biber et al, 2004; Hyland, 2008a; Simpson-Vlach & Ellis, 2010). In doing so, however, the question of how to identify such FS has also been raised. Prior to the availability of computerized corpora, intuition was the primary means of identification. However, within academia at least, this has given way to corpus analyses (Cortes, 2013; Sinclair, 1991). Consequently, lists proposed by both Biber et al (2004) and Hyland (2008a) use frequency of occurrence within a corpus in identifying FS, while Martinez and Schmitt (2012) and Simpson-Vlach and Ellis (2010) choose to temper frequency with other means, including the intuition of NS ‘experts’. While this would suggest that the question of which FS be prioritized for teaching may no longer be an issue, EFL textbook analyses tend to refute this.

As an authoritative source of language input for many L2 learners, textbooks are seen to play an important role (Meunier, 2012; Richards, 1998), and it would seem a reasonable expectation that the language in them be judiciously selected. Koprowski (2005) examined the FS across three different general EFL textbooks aimed at learners of the same proficiency level, assigning a usefulness value to each. He not only found many of those presented to be of little or no value, but also that there was not a single FS shared across the three textbooks, leading him to conclude that intuition may have been the primary means of selection. McAleese (2013) conducted a similar study, finding that, of the FS identified, “a significant proportion... [were] unrepresentative of real-life English” (p.326), suggesting again that recourse to intuition may take precedence over empirically derived data. This is refuted to a degree by Burton (2012), who reports that, of 13 textbook authors surveyed on their use of corpora in selecting lexis, 8 reported having done so. Overall, however, the literature relating to the content of FS within ELT textbooks does not come down in favor of textbook authors, but rather concludes with calls for a more informed selection (Gouverneur, 2008; Harwood, 2014; Meunier, 2012).

A further pedagogical issue with regard to FS relates to their acquisition. While Ellis et al (2008) propose that frequency of occurrence may be the primary factor in

L2 learners' determination of FS, research also suggests that explicit attention to such sequences may be necessary for successful acquisition (AlHassan & Wood, 2015; Peters & Pauwels, 2014; Wood, 2010). AlHassan and Wood (2015, p.53) ascribe this to FS having either so transparent a meaning that they go unnoticed, or, conversely, that their opacity creates such difficulties that they are bypassed. Their study, which examined the effects of focused instruction of FS, concludes by calling on materials writers to provide "noticing and deep processing" (p.61) activities through which learners can develop both recognition and utility with them. Likewise, Peters and Pauwels (2014) comment that L2 learners' use of FS is often not genre-appropriate, and highlight the value of explicit instruction. Unfortunately, in this regard too, textbooks may be failing students, with Wood's (2010) analysis of a number of EAP textbooks' treatment of FS revealing them to be "not particularly effective" (p.103).

## **Research Questions**

In light of the above, relating primarily to the usefulness of the FS explicitly highlighted within a textbook, and the resulting activities focusing on their utilization, this study will aim to answer the following research questions:

1. How useful are the FS given explicit attention within the textbook?
2. To what extent are activities provided which necessitate utilizing these FS?

The textbook selected for analysis is 'Writing Essays: from paragraph to essay' (Zemach & Ghulldu, 2011), the third book in a series of four aimed at developing learners' academic writing skills. This particular textbook in the series is intended for learners at IELTS level 4.5 to 5.5. The reason for selecting this textbook for analysis is that it is used by the author as the basis of a first year undergraduate writing course syllabus at a Japanese university. Also, as noted above, academic writing is repeatedly highlighted as a genre in which FS prevail, with AlHassan and Wood (2015) referring to them as the "building blocks of academic discourse"

(p.52). Furthermore, authors of textbooks focusing on academic writing have available to them various lexical resources, including corpora and lists of genre-specific FS, upon which they may base their selection. It therefore seems reasonable to assume that recourse to such resources be reflected in the FS content of the textbook.

## **Methods**

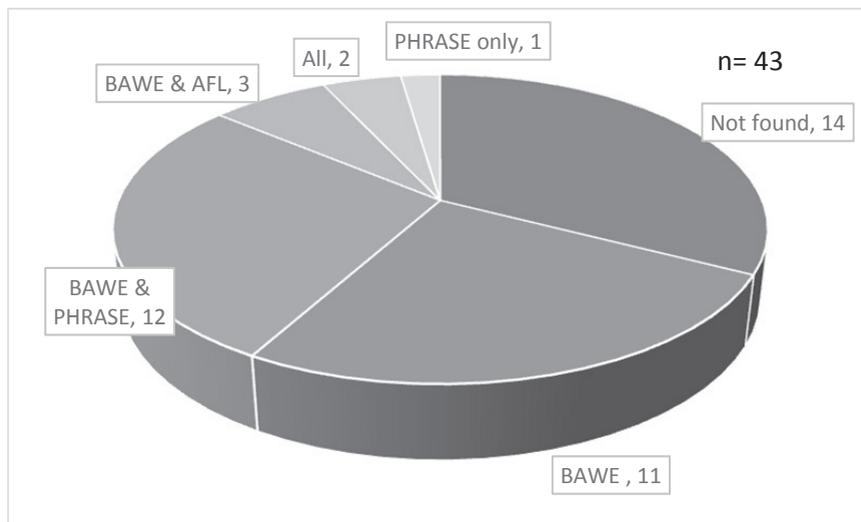
Any sequences of two or more contiguous words explicitly highlighted for attention were extracted by a manual search of the textbook. The usefulness of the extracted sequences was then assessed by determining their presence within the Academic Formulas List (AFL) (Simpson-Vlach and Ellis, 2010) and the Phrasal Expressions List (PHRASE) (Martinez and Schmitt, 2012), and from the incidence of each within the Corpus of British Academic Written English (BAWE) (University of Oxford Text Archive, 2014). With regard to the lists, PHRASE was chosen as it aims to compile the “most frequent formulaic sequences in English” (Martinez and Schmitt, 2012, p.302), indicating how common each is within general spoken, general written, or written academic discourse. AFL was selected as it lists FS “that are significantly more common in academic discourse than in non-academic discourse and which occupy a range of academic genres” (Simpson-Vlach and Ellis, 2010, p.487-488). BAWE was used as a comparison corpus as it comprises over 6.5 million words of text taken from undergraduate and post-graduate student essays. As Writing Essays (Zemach & Ghulldu, 2011) claims to “develop learners’ academic writing skills” (back cover), it seems a fair assumption that any language deemed worthy of explicit attention will surely be useful to this end, and could thus reasonably be expected to be found within BAWE. While the frequency of occurrence used within different studies varies considerably (see Adel & Erman, 2012, p. 82), the cutoff rate of 10 times per million words used by Biber et al (1999), described as “very flexible” (Hsu, 2014, p.148) in comparison to that used in other

studies, appears to be the most generous, and was used herein. Anything occurring fewer than 10 times per million was therefore termed unidiomatic, keeping in mind the genre-specificity of FS.

As with the FS, activities provided to enhance learners' recognition and utilization of FS were also identified through a manual search of the textbook. This aspect of the analysis drew on Peters and Pauwels' (2014) categorization, in which activities are categorized according to one of three functions: recognition, cued output and spontaneous use. Within *Writing Essays* (Zemach & Ghulldu, 2011), recognition exercises included those where learners identify words or phrases that serve a particular function within a text, sentence ordering, and identifying the appropriate one of two given expressions. Cued output exercises included gap-fills, text editing, and vocabulary reviews in which learners are asked to recall phrases. Spontaneous production exercises required learners to write sentences or paragraphs on a given topic or with a particular aim. With a number of such exercises throughout the textbook, only those given as a means of practicing a paragraph style or writing feature within which FS were explicitly highlighted were included. After the initial categorization, activities were then classified as either 'necessary' or 'unnecessary' depending on whether or not use of the FS was required to complete them. For spontaneous output exercises to be classified as 'necessary', explicit reminders to include the target lexis were required.

## **Findings**

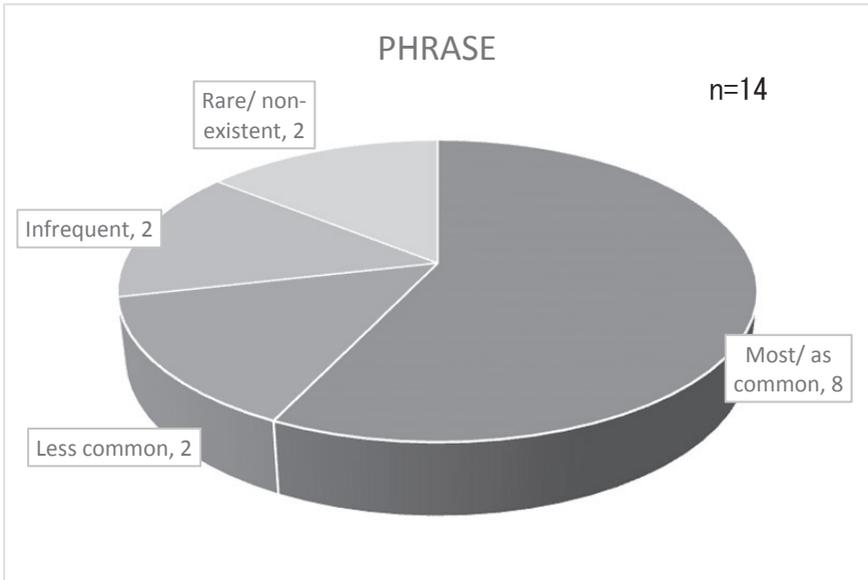
The 43 FS that were identified within the textbook account for just under 32% of the vocabulary items explicitly highlighted (in such sections as FS appeared; vocabulary from sections where FS were not present was not included). Figure 1 shows the occurrence of each of these within both the comparison corpora, BAWE, and the two lists, AFL and PHRASE. It reveals that of the 43 FS found, approximately one third were found to be unidiomatic. Of the 28 deemed idiomatic,



**Figure 1:** The total number of FS extracted from the textbook, as found within BAWE, PHRASE and AFL

17 can be seen to feature in one or other of the lists, with only 2 of these featuring in both, while one of the unidiomatic FS was found to be present in PHRASE. Keeping in mind the low cut-off rate used, one striking factor to emerge from this is the percentage of FS deemed to be unidiomatic when referenced against BAWE, possibly made more so by the fact that 4 of these sequences were not found to occur even once in a corpus of over 65 million words. With a further 8 FS found to occur fewer than 20 times in the entire corpus, it could be argued that using FS the textbook authors propose as useful in academic writing might rather mark them as outsiders to this particular discourse community. The sequence ‘doesn’t have to’ may be seen to exemplify this. While featuring 9 times within BAWE, quick reference to two academic writing textbooks makes it clear that contractions are “not common in academic English” (Bailey, 2011, p.213), or, put more emphatically, “unacceptable” (VanGeyte, 2013, p.41). In contrast, the uncontracted form ‘does not have to’, while still unidiomatic, occurs 29 times in BAWE. The data shown in Figure 1 suggests, therefore, that rather than addressing the difficulties Peters and

Pauwels (2014) highlight regarding learners' recognition of genre-appropriate FS, those highlighted within the textbook may in fact be hindering them.



**Figure 2:** The FS as they occur within PHRASE

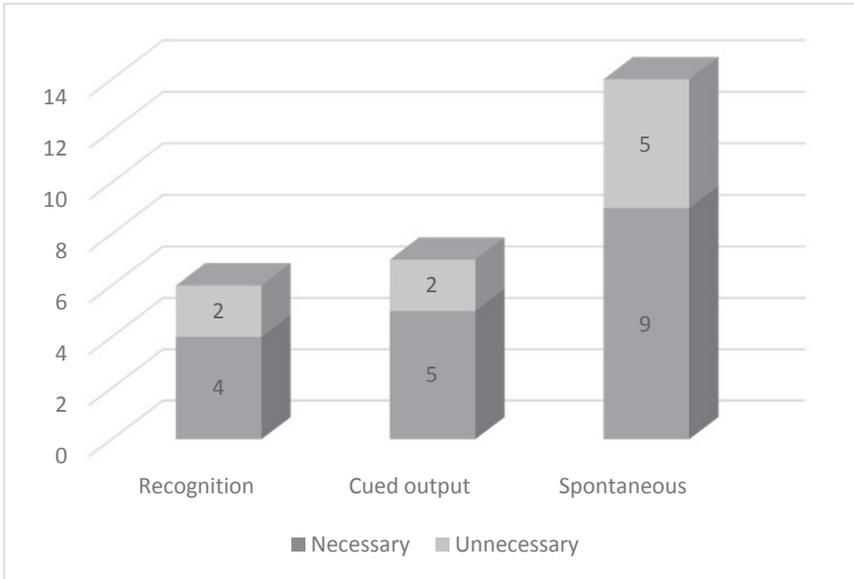
Through a closer examination of the FS from the textbook found to be in PHRASE, it is possible to get further insight into their usefulness in terms of academic writing, as it indicates how likely each sequence is within spoken, written and written academic discourse. Figure 2 shows that, of the 14 FS from the textbook that were identified in PHRASE, just over half are listed as ‘most/ as common’ within written academic discourse. On the whole, these FS occur with considerably greater frequency within BAWE, indicating some level of correlation between the two (see Appendix 1). This does mean, however, that just under half are listed as being either ‘less common’, ‘infrequent’ or ‘rare’, with the single unidiomatic FS featured in PHRASE noted as being ‘infrequent’ in academic writing. Of the other

three FS which are rated 'infrequent' or 'rare', two occur with reasonably low frequency within BAWE, while the remaining one is roughly four times the cut-off rate.

Within AFL, only 5 of the 43 FS extracted from the textbook featured, of which two were also found in PHRASE. Additionally, those found within AFL all had a high rate of occurrence within BAWE. Like PHRASE, AFL also discriminates between genres, the distinction here being between written academic and spoken academic sequences, with those from the textbook all found within the written academic list. While the total of 5 FS found within AFL is considerably less than the 14 within PHRASE, a more accurate comparison would be 5 in the former against 8 in the latter, the remaining 6 in PHRASE being of less significance within academic writing. With 2 FS in common, this amounts to 11 of the 43 FS, or roughly one quarter, found in the textbook seen to be both idiomatic and of particular value within academic writing. Therefore, in seeking to examine the usefulness of the explicitly highlighted FS within the textbook, the data from this part of the analysis suggests that this language aspect would benefit from considerable changes to the authors' methods of selection. It does very little to refute the assertions highlighted above regarding textbook authors' lack of reference to corpora when selecting lexis for inclusion, but rather suggests that such criticism may be well justified.

## **Recognition and Utility**

In terms of recognizing and utilizing the FS, the activities used in the textbook fell within one of the three categories outlined above, namely recognition, cued output and spontaneous use. Figure 3 shows the number of exercises within each of these categories in which using the highlighted FS was either necessary or possible for completion. It can be seen that within each category, the number of exercises where this is the case is roughly double those where it is not. However, this data



**Figure 3** The number of exercises within each of the three categories where using FS was deemed necessary

alone may be slightly misleading. While within recognition exercises, FS were present in almost 31% of instances, roughly proportionate to their incidence within the overall lexis explicitly highlighted, in a number of cases FS constituted only one of the many answers sought, or could be avoided altogether. This was more so within cued output exercises, where the maximum potential use of FS was 47%, assuming they were used in favor of single word items at every possible opportunity. However, should learners instead choose to use single word items wherever possible, and with nothing in the instructions to suggest they do otherwise, this figure dropped to 12%. Within spontaneous output exercises too, while in the majority, reminders encouraging learners to make use of the highlighted lexical items were given in the instructions, whether or not they would include FS is open to conjecture. In accounting for less than one third of the highlighted lexis, in neither output scenario does this seem likely.

An analysis of the recognition and utilization exercises within the textbook does therefore reveal instances in which learners are provided with the opportunity to see FS in context, to produce them within controlled exercises, and to apply them within their own writing. This might suggest that the “noticing and deep processing” AlHassan and Wood (2015, p.61) claim is necessary for successful acquisition could occur. However, that the FS constitute a minority of the lexical items explicitly highlighted, and are given no emphasis over and above the single word items, provides no reason to assume that learners would favor their use in the output exercises. If, as Peters and Pauwels (2014) suggest, learners’ gains are more significant through cued output exercises than recognition, requiring that they be used in more than 12% of instances must surely be necessary. As it is, in terms of aiding learners’ utilization of FS, the treatment afforded this within the textbook arguably falls short, going some way towards confirming Wood’s (2010) somewhat negative evaluation of this aspect in his analyses.

## **Conclusion**

In being seen both to constitute the building blocks of academic writing, and as an indication of ‘belonging’ within the field, the value of FS has been repeatedly highlighted. Alongside this importance, however, the difficulty L2 learners have in successfully acquiring and utilizing them has also been emphasized. Given the genre-specificity of FS, research has aimed to uncover which FS would be most beneficial for L2 learners with regard to academic writing, and also how best to bring about learners’ acquisition and utilization of these. In light of this, the current study has sought to conduct an analysis of the textbook ‘Writing essays: from paragraph to essay’ (Zemach & Ghulldu, 2011) to examine both the usefulness of the FS explicitly highlighted, and how far the textbook goes in accommodating learners’ recognition and command of these. The data from both parts of the analysis reveal that significant changes may be required to better meet the needs of learners.

While the textbook in question is aimed at learners of a slightly lower proficiency level than is required for entry into most UK university undergraduate programmes, given the significance of FS within this discourse genre, and the troubles L2 learners are reported to encounter with them, it would surely be of benefit to begin building recognition and utilization of them at an early stage. The analysis conducted here suggests that neither of these factors is being addressed. Rather, a significant percentage of the FS highlighted for explicit attention within the textbook were found to be unidiomatic, and of those deemed idiomatic, only one quarter were seen to be especially so within academic writing. Furthermore, exercises aimed at improving overall proficiency with FS were found to be lacking, especially those concerned with output, where only a small percentage required the use of FS. However, given the overall value of the FS highlighted for use, this may ultimately be of service to those working through the textbook.

As noted above, many textbook users, both teachers and learners, look to textbooks as providing a rich source of language, and with this must surely come the assumption that this source has some basis in real language use. The results of this analysis, however, suggest otherwise, and can be seen instead to stress the importance of adopting a critical stance to textbook content. Given the relative ease with which corpora, and an increasing number of corpus-informed lists relating to FS within various discourse genres, may be consulted, it is unfortunate that many textbook writers and publishers apparently still do not take the time to do so. Findings from a growing body of work within SLA exploring how L2 learners might best acquire such language must also begin to filter through to textbooks. Otherwise, even with a more judicious selection of language, textbook writers will still be failing learners.

## Appendix 1

Below is listed each of the FS drawn from the textbook, showing where each was found to occur.

The asterisk markings show how apparent each FS is within academic writing within PHRASE.

\*\*\*= most/ as common

\*\*= less common

\*= infrequent

X= rare/ not apparent

Unidiomatic	BAWE	PHRASE (+BAWE)	AFL (+BAWE)	All
the last step	has to	prior to ***	for example 117	as a result *** 22
doesn't have to	must not	at this point ***	such as 172	on the other hand ** 1
not the same as	the same as	in contrast ***	similar to 135	
to meet this	different from	in spite of ***		
one solution is	in order to	in addition ***		
one answer is	compared to	for instance ***		
one thing we can do	in general	in particular ***		
a final answer is	in conclusion	in front of **		
in summary	to conclude	next to *		
to summarise (-ze)	on top of	ought to X		
in back of	in the middle of	just as X		
on the bottom of	after that			
to the right of				
to the left of				
at last (* in PHRASE)				

## References

- Adel, A. & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: a lexical bundles approach. *English for Specific Purposes*, 31, 81-92.
- AlHassan, L. & Wood, D. (2015). The effectiveness of focused instruction of formulaic sequences in augmenting L2 learners' academic writing skills: a quantitative research study. *Journal of English for Academic Purposes*, 17, 51-62.
- Bailey, S. (2011). *Academic writing: a handbook for international students* (third edition). Abingdon: Routledge.
- Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow, England: Pearson.
- Biber, D., Conrad, S. & Cortes, V. (2004). If you look...: lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25-3, 371-405.
- Boers, F. & Lindstromberg, S. (2012). Experimental and intervention studies on formulaic sequences in a second language. *Annual Review of Applied Linguistics*, 32, 83-110.
- Burton, C. (2012). Corpora and coursebooks: destined to be strangers forever? *Corpora*, 7-1, 91-108.
- Conklin, K. & Schmitt, N. (2008). Formulaic sequences: are they processed more quickly than nonformulaic language by native and non-native speakers. *Applied Linguistics*, 29-1, 72-89.
- Cortes, V. (2013). The purpose of this study is to: connecting lexical bundles and moves in research article introductions. *Journal of English for Academic Purposes*, 12, 33-43.
- Ellis, N., Simpson-Vlach, R. & Maynard, C. (2008). Formulaic language in native and second language speakers: psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly*, 42-3, 375-396.

- Erman, B. & Warren, B. (2000). The idiom principle and the open-choice principle. *Text*, 20-1, 29-62.
- Gouverneur, C. (2008). The phraseological patterns of high-frequency verbs in advanced English for general purposes: A corpus-driven approach to EFL textbook analysis. In F. Meunier & S. Granger (Eds.), *Phraseology in foreign language learning and teaching* (pp. 223-246). Amsterdam: John Benjamins.
- Greaves, C. & Warren, M. (2010). What can a corpus tell us about multi-word units? In A. O’Keeffe & M. McCarthy (Eds.), *Routledge handbook of corpus linguistics* (pp. 212-226) [electronic resource]. London: Routledge.
- Harwood, N. (2014). Content, consumption, and production: three levels of textbook research. In N. Harwood (Ed.), *English language teaching textbooks: content, consumption, production* (pp. 1-44). London: Palgrave Macmillan.
- Hsu, W. (2014). The most frequent opaque formulaic sequences in English-medium college textbooks. *System*, 47, 146-161.
- 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*, 18-1, 41- 62.
- Koprowski, M. (2005). Investigating the usefulness of lexical phrases in contemporary coursebooks. *ELT Journal*, 59-4, 322-332.
- Martinez, R. & Schmitt, N. (2012). A phrasal expressions list. *Applied Linguistics*, 33-3, 299-320.
- Mcaleese, P. (2013). Investigating multi-word items in a contemporary ELT course book. In N. Sonda & A. Krause (Eds.), *JALT2012 conference proceedings*. Tokyo: JALT.
- Meunier, F. (2012). Formulaic language and language teaching. *Annual Review of Applied Linguistics*, 32, 111-129.
- Millar, N. (2011). The processing of malformed formulaic language. *Applied Linguistics*, 32-2, 129-148.

- Peters, E. & Pauwels, P. (2014). Learning academic formulaic sequences. *Journal of English for Academic Purposes*, 20, 28-39.
- Richards, J. (1998). Textbooks: help or hindrance? In J. Richards (Ed.), *Beyond training* (pp.125-140). Cambridge: Cambridge University Press.
- Schmitt, N. (2010). *Researching vocabulary: a vocabulary research manual*. Basingstoke: Palgrave Macmillan.
- Simpson-Vlach, R. & Ellis, N. (2010). An academic formulas list: new methods in phraseology research. *Applied Linguistics*, 31-4, 487-512.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- University of Oxford Text Archive, (2014). *British Academic Written English Corpus*. Retrieved August 15, 2015, from <http://ota.ahds.ac.uk/desc/2539>
- Van Geyte, E. (2013). *Writing: learn to write better academic essays*. London: HarperCollins.
- Wood, D. (2010). Lexical clusters in an EAP textbook corpus. In D. Wood (Ed.), *Perspectives on formulaic language: acquisition and communication* (pp. 88-106). London: Continuum.
- Wray, A. (2000). Formulaic sequences in second language teaching: principle and practice. *Applied Linguistics*, 21-4, 463-489.
- Wray, A. (2002). *Formulaic language and the lexicon* [electronic resource]. Cambridge: Cambridge University Press.
- Wray, A. (2009). Future directions in formulaic language research. *Journal of Foreign Languages*, 32-6, 2-17.
- Zemach, D. & Ghulldu, L. (2011). *Writing essays: from paragraph to essay*. London: Macmillan Publishers Limited.

# EFLアカデミック・ライティングの教科書における定型表現と定型表現習得用練習問題の有用性に関する分析

ルイス・マリー

## 要旨

定型表現が第二言語学習者にとって重要であることはしばしば強調されるが、また同時に、定型表現習得の難しさも頻繁に指摘される場所である。本稿は、英語ライティングの教科書 *Writing essays: from paragraph to essay* (『エッセーを書く：パラグラフからエッセーへ』) を分析対象とし、同書が特に着目する定型表現と、その習得のために挙げられた練習問題とが、どの程度有用であるかを検討する。前者のデータについては、コーパス比較を行い、実証に基づいて作成された定型表現リストと対照させた。その結果、同書で取りあげられた定型表現の大部分は、重要度が高いとは言えないことがわかった。また練習問題については、語彙関連の練習問題の大部分が、同書が挙げる定型表現を用いずとも解けるものであるため、定型表現の習得にはあまり役立たないであろうことが示唆される。