Journal of English for Academic Purposes 43 (2020) 100810

Contents lists available at ScienceDirect

Journal of English for Academic Purposes

journal homepage: www.elsevier.com/locate/jeap

The bottom line: Are idioms used in English academic speech and writing?

Julia Miller

School of Education, The University of Adelaide, 10 Pulteney Street, Adelaide, SA, 5005, Australia

ARTICLE INFO

Article history:
Received 17 October 2018
Received in revised form 24 October 2019
Accepted 29 October 2019
Available online 31 October 2019

ABSTRACT

Many researchers with English as an additional language (EAL), as well as EAL teachers and students, believe that idioms are not used in English academic speech or writing. By 'idiom', they usually mean an expression that is figurative, opaque, multiword, largely fixed, and institutionalised, in keeping with traditional EAL textbook definitions. By contrast, researchers addressing multiword expressions in academic corpora have frequently used the term 'idiom' very broadly, resulting in lists that include lexical bundles and other types of collocation, rather than more traditional idioms.

This study identifies all the traditional, textbook-type idioms in the British Academic Spoken English (BASE) corpus, examines the range of texts in which they appear, then compares their frequency to the same idioms used in the Oxford Corpus of Academic English (OCAE). The resulting list of 545 idioms, 56 of which appear in four or more BASE texts and 43 of which appear at least 100 times in OCAE, can be used confidently by EAL writers and teachers to enrich their own and their students' English academic speech and writing.

1. Introduction

It is often assumed by language teachers and learners that idioms are too informal to appear in academic English. A scrutiny of six advanced level EAP books, for example, reveals that only one (de Chazal & Moore, 2013, p. 150) includes even half a page on idioms, and O'Dell and McCarthy's comprehensive *English idioms in use: Advanced* (2010) self-study book contains only one section on idioms that can be used in essay writing. Researchers and students with English as an L2 may therefore avoid using idioms in English academic discourse. Idioms, however, perform many functions in academic communication: describing and evaluating, emphasising, paraphrasing, creating a sense of group identity, and marking a change of topic (Simpson & Mendis, 2003). Deliberate exclusion of idioms from one's writing may therefore signal a lack of phraseological competency, which can mark a writer out as uninformed of the conventions of a discourse community (Li & Schmitt, 2009, p. 86). Even more importantly, those who misinterpret or fail to understand idioms may not only miss essential information (Martinez & Murphy, 2011), but misunderstand central concepts entirely (Liu, 2003, p. 687), especially if their knowledge of the idiom's individual words gives them a false confidence in their understanding of the whole text (Park & Chon, 2018). Phraseological understanding is “particularly relevant” to advanced learners and users of English (Kremmel, Brunfaut, & Alderson, 2017, p. 865), such as L2 university lecturers and students. Due to their own misconceptions or their teachers' misgivings, however, L2 writers and students may be unaware of idioms that would help them understand English

E-mail address: julia.miller@adelaide.edu.au.

https://doi.org/10.1016/j.jeap.2019.100810
1475-1585/

academic texts or lectures, or enrich their own academic speaking and writing, and are therefore not confident which idioms might be safe to use in their own academic work. Even those instructors who do teach idioms generally do not know which expressions are used most often in academic environments (Martinez, 2013, p. 197), and there is a need for more materials targeted at teaching phraseological content (Kremmel et al., 2017).

Materials for academic idiom teaching should, nevertheless, consider the register in which they appear (Liu, 2008). Biber (2006, pp. 222–223) cautions that "all spoken university registers, regardless of purpose, are sharply distinguished from written university registers in most of their typical linguistic characteristics". In regard to idioms, it is therefore appropriate to follow Liu's advice (2008, p. 113) and examine "[how] formal an idiom is and what register and language variety it is used in [as these] constitute perhaps the most important information for language learners in order for them to use it appropriately". For this reason, a comparison of spoken and written corpora can highlight the relative frequency of idioms in each register.

This paper therefore centres on the use of idioms (e.g. *the bottom line*) recorded in the British Academic Spoken English corpus, compared with the same idioms used in written academic texts from the Oxford Corpus of Academic English. The aim of the research was to discover whether idioms are actually used in academic English and, if so, which are used more in speech or in writing. The findings have important implications for the speech and writing of English L2 university students and lecturers, since any frequently used idioms can be confidently used in their own writing and presentations, and less frequent items can be noted for decoding purposes.

2. Definition of the term 'idiom'

The lack of consensus over terms used in phraseological research is a cause of great confusion. The study of phrases was well advanced in eastern Europe before researchers elsewhere became aware of its existence, resulting in a plethora of terms for the same type of expression confounded by differences in use of some of those terms. One person's 'idiom' may therefore be another person's 'phraseological unit'. An overview of the prevailing terms from a phraseological perspective is given in Miller (forthcoming).

In brief, words collocate on a continuum from free combinations (e.g. *an interesting question*); to restricted collocations (e.g. *a thorny question*); to figurative idioms, whose meaning is transparent and can be deduced from the literal meaning (e.g. *ring a bell*); to pure idioms, which are opaque to those unfamiliar with them (e.g. *red herring*) (Howarth, 1998, p. 164). The difficulty in definition lies in determining both where the term 'idiom' fits on this continuum, and which term to use. Common terms are "figurative idioms" (Cowie, 1998; Howarth, 1996), "pure idioms" (Cowie, 1998; Howarth, 1996; Moon, 1998), and "idioms" (Dobrovol'skij & Piirainen, 2005; Simpson & Mendis, 2003). All these researchers define idioms as institutionalized (commonly accepted by other first language (L1) speakers in the speech community); reasonably fixed syntactically; composed of more than one word; and figurative, with varying degrees of opacity (sometimes also called 'compositionality'). This paper will adopt this definition of the word 'idiom', which covers both idioms in the everyday understanding of the word, exemplified by the online *Macmillan dictionary* (e.g. *to have your feet on the ground*), and more common but nevertheless figurative expressions (e.g. *on one hand*).

Phrasal verbs (e.g. *turn up*) are not included in the study, owing to their ubiquity and the need, as in other studies (e.g. Moon, 1998), to limit the scope of the research. The term 'multiword expression' (MWE) will be used for all other phraseological expressions, regardless of whether the original writer has used the term 'idiom'.

Idioms in this narrower sense have been chosen as the focus of this paper because they are under-researched in academic English (only Simpson and Mendis (2003) have focussed on this category in academic speech) and yet may form an important part of academic speech and writing, since failure to comprehend or correctly use them can inhibit L2 users' integration into the academic community. Previous studies have addressed the functions of MWEs in academic English, in MICASE (Simpson & Mendis, 2003); in an Academic Formulas List (Simpson-Vlach & Ellis, 2010); and in student writing (Shin, 2019). The current study aimed to establish the presence of idioms by English L1 lecturers and in published academic writing in English, facilitating future research on the function of these idioms. Past research has focussed mainly on spoken corpora. So far, no comparison has been made between idioms used in spoken and written academic English.

Research on phraseology and academic English

Previous studies of academic English have examined a range of MWEs by a variety of means and in a variety of corpora, using frequency and, often, intuition, to identify expressions useful for L2 learners.

One of the earliest studies on idioms in English academic use is that of Simpson and Mendis (2003), who devised a list of idioms in the Michigan Corpus of Academic Spoken English (MICASE). MICASE (recorded 1997–2001) contains 1,848,364 words in 152 transcripts from a wide range of disciplines. The transcripts are taken not only from lecture and seminar recordings, but also from individual consultations and small study group sessions, meaning that student voices are included as well as those of staff. Simpson and Mendis defined an 'idiom' as a multiword, fixed, institutionalized, semantically opaque expression, and read "at least half the transcripts" in MICASE to identify idiom candidates, using "the collective intuition of three raters" (2003, p. 424). They then checked the frequency of these idioms in MICASE before examining the idioms' pragmatic functions in the corpus. Their final list includes 238 different idioms, only 32 (10%) of which appear more than four times in MICASE. They conclude that "idioms are neither rare nor particularly frequent in academic speech", with a frequency of 260 tokens per million words when repetitions in the same speech event are discounted (2003, p. 427). Unfortunately,

Simpson and Mendis do not report on the range (that is, dispersion across transcripts) of the idioms in their findings. For example, their most frequent idiom, *bottom line*, is reported as occurring 17 times, but no mention is made of the fact that it occurs in only 12 transcripts, with one occurrence possibly a literal, non-figurative use of the expression. Like other researchers, they admit the difficulties of defining an 'idiom', and their criteria do not necessarily match those of other studies. For example, *making out like bandits* is listed as a transparent idiom, but the meaning may not be understood outside the United States. Simpson and Mendis found no large differences in frequencies between the humanities and the sciences, and point to the frequency and pragmatic functions of their idioms as evidence that they should be included in teaching materials of English for academic purposes (EAP).

A similar study is that of Liu (2003), which identified multiword expressions in three spoken American English corpora: Barlow's Corpus of Spoken, Professional American English; his own American media English corpus; and MICASE (Liu, 2003, p. 677). Liu adopted Fernando's (1996) classification of idioms as pure (e.g. *kick the bucket*), semiliteral (e.g. *go through*), and literal (e.g. *according to, throw away*), and included phrasal verbs in his study, identifying 9683 MWEs that appeared in at least two of four idioms dictionaries or two of three phrasal verb dictionaries. He then used MonoConc Pro 2.00 software to search for these MWEs in two of his corpora, and the MICASE website search facility to locate them in MICASE. Liu reports on differences in register, variations of form (*in the long run/term*), and preferred tense. The study is excellent, but Fernando's (1996) classification of 'idiom' is extremely broad, including literal expressions and acronyms. Among her examples, for instance, is the adjective *blue* meaning 'obscene'. Thus, "*blue film/joke/gag/story/comedian*" are regarded as lexical variations of an idiom (Fernando, 1996, p. 36), rather than collocations of an adjective and a noun. Her literal idioms include expressions such as *tall, dark and handsome* and *a* (*very*) *happy birthday*, termed 'formulae' by other researchers (e.g. Moon, 1998). Liu's criteria for 'idioms' in his study are thus very wide. In fact, many of the MWEs in Liu's study could be termed 'lexical bundles', which Biber (2006, p. 134) defines as "the most frequently occurring sequences of words". Liu's 302 most frequent MWEs have a rate of at least 2 per million words (pmw), with 47 occurring as frequently as 50 pmw. These he refers to as "frequently used" (2003, p. 680).

Grant and Bauer's research (2004) is narrower in focus than Liu's study and concentrates on "core idioms" as a set of particularly problematic MWEs that learners need to know. The authors separate core idioms from figurative expressions, explaining that a core idiom should be a multi-word, non-compositional (i.e. opaque), non-figurative expression. In their description, "'figuratives' can be made sense of. Idioms cannot be" (2004, p. 52). They then give a test that can be applied to MWEs to see if they meet the criteria for core idioms, and recommend that these core idioms be taught first to learners. Grant (2007) later compared figurative idioms in spoken subcorpora from the British National Corpus (BNC), Liu's contemporary spoken American English corpus (2003), and MICASE, in order to establish a frequency list for teaching purposes. Grant's criteria are narrower than Liu's and do not include "phrasal verbs (deal with, go through, find out, etc.), 'vague language' (kind of, sort of), fixed expressions (in fact, at all, in order to/that, etc.), and discourse markers (first of all, according to, etc.)" (Grant, 2007, p. 171). She found a total of 76 "figuratives", only 10 of which appeared with a rate higher than her frequency threshold of 2 pmw.

Other studies on phraseology in academic speech and writing have focussed on units that are closer to collocations than idioms. For example, Simpson-Vlach and Ellis (2010)'s Academic Formulas List uses Coxhead's Academic Word List (2000) to identify "frequent recurrent patterns in written and spoken corpora that are significantly more common in academic discourse than in non-academic discourse and which occupy a range of academic genres" (Simpson-Vlach & Ellis, 2010, pp. 487–488). These are listed according to the authors' metric of "formula teaching worth" (2010, p. 488). Although they started their list by measuring frequency, they then applied a "mutual information" measure (2010, p. 490), which indicates the degree to which words are likely to co-occur. The corpora in their study were MICASE, British National Corpus academic speech and selected academic writing files, and a 1.2 million word corpus of research articles created by Hyland in 2004. Twenty EAP language testing experts and instructors were also asked to judge whether they felt the formulae were worth teaching**.** The metric of formula teaching worth is thus derived from frequency data but confirmed partly by intuition. Their frequency cut-off threshold was established at 10 per million words.

Liu (2012) later built on Simpson-Vlach and Ellis's (2010) work by examining the frequency and use of "multi-word constructions", "including LBs [lexical bundles], idioms, and phrasal/prepositional verbs" (Liu, 2012, p. 28) in academic English writing, by means of written academic sub-corpora in the BNC and the Corpus of Contemporary American English (COCA). Liu's expressions were chosen from lists constructed by previous researchers (such as Biber & Conrad, 1999, and Simpson-Vlach & Ellis, 2010) and idiom or phrasal verb dictionaries, selecting items partly by intuition, and there are very few traditional idioms (such as *the bottom line*) in his final list. He makes the point, however, that lack of frequency does not equal lack of importance to learners, as students may miss information in a lecture or reading because an opaque idiom is used. Liu recommends a frequency of 20 pmw, together with range criteria, to establish an MWE as 'frequently used'.

Another study based on frequency is Martinez and Schmitt's work on phrasal expressions, defined as "fixed or semi-fixed sequence[s] of two or more co-occurring but not necessarily contiguous words with a cohesive meaning or function that is not easily discernible by decoding the individual words alone" (2012, p. 304). To be included in the list, an expression must be a "morpheme equivalent unit" (Martinez & Schmitt, 2012, p. 308 following Wray, 2008) which is processed as a whole. In addition, it must be semantically opaque or "deceptively transparent" (Martinez & Schmitt, 2012, p. 313). For example, a learner might think they understand an expression such as *for some time*, but they might misunderstand it to mean "a short amount of time" (Martinez & Schmitt, 2012, p. 309). Martinez and Schmitt (2012, pp. 309–310) give three "auxiliary criteria": there may be a one word equivalent of the expression, in English or another language; literal translation into a learner's L1

could affect their understanding of the expression; and a change in grammar could lead to misinterpretation of the expression. Here, they give the example "as a host *I'm expected to* be courteous" (2012, p. 310), where the passivization of the verb *expect* changes the meaning of *expect* from *thinking something will happen* to *supposed to*. Martinez and Schmitt (2012) used *WordSmith Tools* to extract n-grams (words that co-occur) of between two and four words that occurred at least 787 times in the BNC and then read through the n-gram list using their six selection criteria to identify items for their phrasal expressions list. One problem they encountered was "phraseological polysemy" (2012, p. 311), where a phrase has several different meanings. To counter this, they examined "a random concordance sample of 100 lines" to find a percentage of occurrences for the phrase, using a further random sample as a comparison in order to achieve a consistent result (2012, p. 312). The final phrase list has 505 expressions, which are "almost entirely composed of the top 2,000 words in English, with the vast majority in the top 1,000" (2012, p. 313). The most frequent phrase on their list is *have to* (expressing obligation), and the last on their list is *come about*. Thus, while all their phrases are to some extent opaque, none could be classified as an idiom according to the definition used in my study.

Martinez returned to the idea of frequency in 2013, qualifying it by saying that "just because an item does not appear frequently in a corpus should not necessarily indicate that it is not useful" (2013, p. 187) and developing a Frequency-Transparency Framework (2013, p. 190). From this framework, he recommends that highly frequent expressions should be prioritised in teaching, but that where opaque and transparent expressions appear with equal frequency, the more opaque expressions should be given teaching priority over the transparent items. He includes collocations (*guide decisions, explicit instruction*) as transparent, compared to more opaque idioms such as *black and white* and *hard and fast*. He cautions, however, that 'frequent' is a relative concept and adds that "one could also use good sense and judgement and scan the list for when words begin to seem less 'useful' "(Martinez, 2013, p. 194). Frequency is thus used as a starting point, but complemented by subjective judgement.

Range is another important indicator of an idiom's usefulness. 'Range' in this case refers to an idiom's dispersion across a number of "different texts and/or academic disciplines and divisions" (Liu, 2012, p. 26). Liu (2012) chose expressions with 20 tokens per million words and followed Simpson-Vlach and Ellis (2010) in setting a range of six of eight disciplines in COCA (one of which had to be in science or engineering) or five of six in the BNC. In other words, each expression had to appear in six of eight COCA disciplines or five of six BNC disciplines. The greater an idiom's dispersion across disciplines, the more useful it will be for learners' understanding of its use in academic spoken and written genres generally, and an examination of its use in particular contexts will help them to understand more about the practices of individual disciplines (Groom, 2005).

Even when range and frequency have been considered, most of the aforementioned writers (e.g. Liu, 2012; Martinez, 2013; Simpson & Mendis, 2003; Simpson-Vlach & Ellis, 2010) highlight the subjectivity involved in choosing which idioms will be the most useful. Liu (2003, p. 672) initially criticised intuition as the basis for the inclusion of idioms in English teaching materials, arguing that such intuition can result in the selection of little used and, therefore, "unhelpful" idioms. In later work, however, Liu (2012) used intuition to support decisions based initially on frequency. Simpson-Vlach and Ellis similarly highlight the "intuitive weeding of purely frequency-based lists… [as] methodologically tricky and open to claims of subjectivity" (2010, p. 490), although Simpson and Mendis (2003, p. 435) list 20 idioms they "believe" to be useful. Martinez (2013), however, indicates the need for personal judgement. Whatever the method used, a starting-point is needed in determining which idioms are useful, even after a frequency count has been made.

There is also debate in the literature over whether to present idioms for general or discipline-specific use. Hyland and Tse (2007) maintain that a generic wordlist will not apply across disciplines, and Durrant (2009) cautions that the vocabulary needs of students vary according to discipline. Ackermann and Chen (2013), however, point out that EAP lessons are generally conducted cross-disciplinarily, and a generic list will have benefits for all students. Liu (2012) also argues that general academic English is useful to learners, as they are more likely to need general academic writing skills before they need to gain discipline-specific skills.

This study will therefore identify idioms, their frequency and range in a cross-disciplinary academic spoken corpus (British Academic Spoken English, or BASE), before examining the frequency of these idioms in a cross-disciplinary academic written corpus (the Oxford Corpus of Academic English, or OCAE). L2 writers, students and teachers will then have a list of the idioms most frequently used in both spoken and written British academic English, promoting greater access to the English academic speaking and writing community as a whole.

4. Method

As stated earlier, this paper uses the word 'idiom' to include multiword expressions (from two words to a sentence in length) that are institutionalized, reasonably fixed syntactically, figurative, and, to a greater or lesser extent, opaque. The search for these idioms was conducted in two corpora—BASE and OCAE—using Sketch Engine software.

Sketch Engine uses the term 'token' to refer to "the smallest unit that each corpus divides into" (Lexical Computing CZ, n.d.), so that punctuation marks and word forms are each counted as tokens. A phrase can also be counted as a token in Sketch Engine. The word 'token' is therefore used in this study to mean 'an instance of an idiom' and comparisons are made 'per million words' (pmw) which is how Sketch Engine reports token frequency. The notion of idiom frequency is, as Liu (2003) notes, fairly arbitrary. I have therefore reported in detail on those idioms with 4 or more occurrences (at least 2.40 pmw) in BASE and 100 or more occurrences (at least 1.18 pmw) in OCAE.

BASE was recorded between 2000 and 2005 at the universities of Warwick and Reading in the UK. It contains the transcripts of 160 lectures and 39 seminars from a wide range of disciplines, each text representing a different subject. The transcripts are grouped under four faculties (or "broad disciplinary groups" (Nesi & Thompson, 2006, p. 1)): Arts and Humanities, Life and Medical Sciences, Social Sciences and Physical Sciences. During my analysis, one Life and Medical Sciences seminar (lssem003) was found to contain content identical to six lectures (lslct020-lslct025) and was therefore not included in the study, giving a final total of 198 transcripts and 1,669,105 tokens. (Sketch Engine does not include the BASE seminar materials in its online version of the corpus, so this number was obtained by uploading the seminar corpus to Sketch Engine and combining the token count for lectures and seminars based on the Sketch Engine tokenising system.) OCAE has 84,423,644 tokens taken from 8,054 documents covering 26 academic disciplines. I have used Sketch Engine's basic measurement of number of tokens per million words to provide comparisons.

Identification of idioms in a corpus is a lengthy process. It cannot be done automatically, since idioms may take many grammatical forms and vary in length. While known idioms can be easily identified in a corpus, this study's aim was to find all the idioms in one corpus (BASE) in order to provide a basic working list, and then to examine their frequency in written academic English via OCAE. For this reason, idioms that were frequent in OCAE may have been missed, but this was unavoidable due to the impossibility of reading the whole of OCAE. Such problems of scale have meant that previous studies have analysed representative samples from a corpus. For example, Simpson and Mendis (2003, p. 424) read around half the transcripts in MICASE. The present study differs in that all the texts in the BASE corpus were scrutinised for idioms. A human approach was used, similar to Ackermann and Chen (2013) and Simpson-Vlach and Ellis (2010). One academic who is an EAP lecturer with a PhD on phraseology established a list of criteria for inclusion using the definition of idioms given earlier in this paper: institutionalisation, relative fixedness, multiword nature and opacity. She and another rater (a university educated L1 speaker of English who is an editor and proof-reader with extensive experience in English academic writing) then rated two texts together to confirm the criteria, before rating 12 texts (three from each of the four faculties in BASE) independently to establish consistency of rating. Any differences were discussed until a consensus was reached. Each rater then independently read and noted the idioms in each of the 198 BASE texts. Once all the potential idioms had been identified, the results from the two raters were checked by an international phraseology expert, Professor Dmitrji Dobrovol'skij, using the same criteria, to increase reliability by verifying that all the idioms identified met the inclusion criteria, and MWEs that still failed to meet the study's idiom definition criteria were removed.

The frequencies of the final list of idioms were verified in BASE using Sketch Engine software. Unlike Simpson and Mendis (2003), I noted the number of separate BASE texts in which an idiom occurred, in order to calculate its textual range. I also excluded all seemingly literal uses of possibly idiomatic expressions in the BASE corpus but, when there were more than 100 tokens in OCAE, I was unable to read them all. For instance, *gold standard*, which can be a technical term or an idiom, appears 562 times in OCAE and 9 times in BASE (3 lectures and 3 seminars, with one speaker using the idiom four times). Two of the BASE tokens are technical (referring to monetary policy) and were excluded as non-idioms, but I was unable to check all the OCAE results manually. Among the first 100 tokens in OCAE, 40 tokens are from economics papers, where one would of course expect to find many technical instances of the expression. By contrast, the first page lists 15 tokens in the field of medicine, all of which are figurative. This makes the method of extrapolating data from example pages potentially unreliable, depending on how the range of texts is organised in the corpus. Nevertheless, the appearance of many figurative tokens within the first five pages, and the large overall number of tokens, indicate that the idiom is used fairly frequently in written texts, and a further search might reveal that many of these are in the medical domain.

Where tokens were removed from my corpus frequency counts, the new frequency was calculated by dividing the number of tokens in idiomatic use by the total number of tokens in the corpus and multiplying by a million. For example, *gold standard* appears figuratively 9 times in BASE, so 9 was divided by 1,669,105 (the total number of BASE tokens) and multiplied by a million, resulting in a frequency count of 5.39 pmw.

One advantage of Sketch Engine is that both BASE and OCAE are lemmatised. This means that a search for *take*, for example, yields all the forms of the verb. The wildcard character * can also be used to search in Sketch Engine, making it possible to find variations, as in *take * for granted*. Some variations could not be predicted, but as these were unlikely to have appeared more than once this is not a concern. For example, in BASE we find *get (some kind of) handle (more generally) on*. I used *a handle on* for the corresponding search in OCAE. Other idioms appeared in an incomplete form. For instance, *on the other* references but does not include the final word *hand*. It was hard to check the frequency in OCAE of incomplete idioms of this kind, as the words could be used in many ways outside the idiom. In such cases, I did a collocation co-occurrence search for nouns (such as *side*) likely to indicate a non-idiomatic expression and subtracted this number from the total to get an estimated number of tokens.

Almost all the speakers in BASE were English L1 speakers (Nesi, 2015) and most were lecturers. I disregarded text spoken by students (identifiable in BASE by the character 's'), so that the study would represent academics' language as far as possible.

5. Frequency of idioms in BASE and OCAE

The total number of idiom uses in BASE, including repetitions and slight variations, was 1393, or 835 pmw. The total number of individual idioms found in BASE, excluding all repetitions, was 545, or 326 pmw. This is greater than Simpson and Mendis's (2003) finding of 260 idioms pmw in MICASE. Moreover, Simpson and Mendis found a total of only 32 idioms that appeared four or more *times* in MICASE. By contrast, I found 55 idioms that appeared in four or more *texts* in BASE and 74

Table 1
Idioms appearing in at least 4 BASE texts and at least 2.00 per million words in OCAE (presented in order of BASE frequency).

Idiom	BASE frequency per million words	Number of BASE texts in which the idiom occurs	Number of faculties in which the idiom occurs	OCAE frequency per million words
bear in mind	46.73	42	4	10.17
on the one hand	37.74	30	4	31.87
balance of power	10.78	8	2	6.19
the bottom line	8.39	8	4	2.50
(take) a step back/ further	7.19	12	4	4.76
bad news	5.39	8	4	2.75
on the face of it	5.39	8	4	2.57
come into play	5.39	8	3	4.26
in the early days	3.59	5	4	2.33
go hand in hand with	3.59	5	3	2.68

idioms that appeared four or more *times* (sometimes appearing more than once in a single text). Forty-three of these appeared in at least 3 of the 4 faculties represented in the corpus. The average number of idioms per transcript in BASE was 7.03. All the frequencies in Tables 1 to 3 are reported per million words (pmw), and all idioms are reported in their canonical form. Table 1 presents the idioms appearing in at least four BASE texts and with a frequency rate of at least 2.00 per million words in OCAE. It does not, however, include those that are more frequent in OCAE than BASE. These are listed in Table 2. Appendix A (online) lists all idioms which appear in at least 2 BASE texts, while Appendix B (online) lists all idioms which appear with a frequency of at least 1.18 per million words (100 tokens) in OCAE.

Expressions in Table 1 such as *bear in mind* and *on the one hand* might be expected to appear in academic lectures, due to their general frequency in speech. The use of *bad news* and *come into play*, however, is less evident for L2 speakers, especially in academic writing. *Bear (something) in mind* was actually the idiom in BASE with the greatest range and frequency, appearing 78 times in 42 different texts, with an overall frequency of 46.73. Its use was greatest in the Life Sciences, where it appeared in 17 different lectures and seminars. It was the sixth most common idiom in OCAE, with a frequency of 10.17. By contrast, *keep in mind* appeared in only one BASE text but 489 OCAE texts (5.79 pmw).

One idiom, *on the one hand*, had a frequency of 37.74 pmw in BASE, with some speakers using it repeatedly. One lecturer used it 8 times in one lecture, while another used it 6 times, two used it 5 times, and several used it 2, 3 or 4 times. It was also used frequently in OCAE (31.87 pmw). The omission of the definite article (*on one hand*) was rare in BASE, occurring in only 3 texts (1.82 pmw). In each case, the speaker also used the phrase with the definite article included. By contrast, *on one hand* appeared 218 times in OCAE (2.58 pmw).

Balance of power was restricted to politics and history lectures in BASE, but the other seven idioms in Table 1 were spread over a range of disciplines in different faculties and appeared with a frequency of at least 2.14 pmw (181 tokens) in OCAE, indicating that they are used in both academic speech and writing.

Table 2 presents idioms with a frequency of at least 2.00 per million words in OCAE that appear more frequently in OCAE than BASE, indicating a greater frequency of use in writing than speech.

The range of texts in which idioms appeared in OCAE was not examined in this study, due to lack of resources. The frequency of tokens is, however, a starting point for discussion, and eleven idioms appearing in hundreds of OCAE texts were used with greater frequency in writing than in speech. This does not preclude their verbal use, but it does indicate that they are often used in written texts. The most frequent idiom in OCAE was *on the other hand*, appearing 7439 times (88.12 pmw),

Table 2
Idioms appearing more frequently in OCAE than BASE, with a frequency of at least 2.00 per million words in OCAE (presented in order of OCAE frequency).

Idiom	OCAE frequency per million words	BASE frequency per million words	Number of BASE texts in which the idiom occurs	Number of faculties in which the idiom occurs
on the other hand	88.12	64.11	30	4
in the light of	34.99	5.39	8	3
on the other [hand]	20.34	10.19	17	3
in the hands of	12.54	6.59	11	3
in its own right	9.27	5.99	10	4
along the lines of	9.24	6.59	9	3
in the long run	7.84	5.39	9	4
gold standard	6.66	5.39	6	1
driving force	6.21	5.39	8	4
last resort	4.05	3.00	5	2
keep in mind	5.79	0.01	1	1

Table 3
Idioms appearing in at least 4 BASE texts but less than 2.00 per million words in OCAE (presented in order of BASE frequency).

Idiom	BASE frequency per million words	Number of BASE texts in which the idiom occurs	Number of faculties in which the idiom occurs	OCAE frequency per million words
at the end of the day	14.98	13	4	1.36
take on board	7.79	13	4	0.00
by and large	7.19	12	4	0.04
take for granted	7.19	9	3	0.01
across the board	5.99	7	3	1.21
at the back of one's mind	5.99	4	2	0.30
what on earth	4.79	7	3	0.17
go without saying	4.79	6	2	1.85
trial and error	4.79	5	3	1.85
down the line	4.19	7	3	0.08
over the top	4.19	6	3	0.05
state of the art	4.19	4	2	1.36
from scratch	3.59	6	3	1.86
bridge the gap	3.59	6	3	1.85
big picture	3.59	6	3	1.30
get one's head (a) round	3.59	5	4	0.02
keep an eye on	3.59	5	3	0.78
hang on a minute	3.59	5	2	0.04
on the spot	3.59	4	4	0.84
get to grips with	3.59	4	3	1.78
go through the roof	3.59	4	2	0.06
full circle	3.00	5	3	0.54
how on earth	3.00	5	3	0.08
that's another story	3.00	5	3	0.27
cast one's mind back	3.00	5	3	0.04
the other side of the coin	3.00	5	2	0.54
ring a bell	3.00	5	1	0.15
the good old days	3.00	4	3	0.27
a grey area	3.00	4	2	0.60
get the picture	2.40	4	3	0.08
spring to mind	2.40	4	3	0.31
the whole story	2.40	4	3	1.41
on the right track	2.40	4	3	0.28
have a stab at	2.40	4	3	0.25
the high point	2.40	4	2	1.52
it's early days	2.40	4	2	0.01

which was over twice the frequency of *on the one hand* (31.87). When making a contrast, therefore, it is not always necessary to use the first part of this often paired expression. In fact, of all the idioms identified in BASE, *on the other hand* was the most frequent in OCAE, pointing to the importance of this phrase in written academic English.

Table 3 presents idioms in at least four BASE texts but with a rate of less than 2.00 per million words in OCAE, indicating a greater frequency of use in speech than writing.

The idioms in Table 3 are reflective more of academic speech than writing, and it is recommended that L2 speakers use them in oral presentations or lectures rather than in academic papers. Even then, some, such as *what on earth* and *hang on a minute*, are more colloquial than others. Since formality is hard to gauge, and not always indicated in English learner's dictionaries, it is recommended that L2 speakers listen for these expressions in academic contexts to learn more about their use before employing them themselves.

6. Discussion

Of the four faculties in BASE, that with the greatest idiom use was Social Sciences (390 idioms, or 233.66 pmw). Arts and Humanities followed, with 318 idioms (190.52 pmw). Life and Medical Sciences was third (304 idioms, or 182.73 pmw), and Physical Sciences had a much smaller number (127 idioms, or 76.09 pmw). In fact, Physical Sciences had no idioms that were not represented in the other faculties. These are figures for the total numbers of idioms used, including repetitions. The findings contrast with those of Simpson and Mendis (2003), who did not find a preponderance of idioms in any one faculty. The most frequent BASE idioms were, however, found in all the faculties, as well as in OCAE. Therefore, despite debate in EAP literature (e.g. Durrant, 2009; Hyland & Tse, 2007) about the value of generic wordlists, the most frequent idioms in this study would benefit L2 students and researchers in all disciplines, as Liu (2012) and Ackermann and Chen (2013) suggest.

That does not mean that all native speakers employ idioms, however, as use varied greatly among the BASE lecturers. Eight BASE transcripts contained 20 or more idioms. One of these was from Arts and Humanities; 3 were from Life and Medical Sciences; and 4 were from Social Sciences. In all cases, each speaker used at least one idiom more than once. For example, although *to sit on the fence* was used in only 2 BASE lectures, an Applied Linguistics lecturer used variations of the idiom 9 times in one lecture, including the truncated form *fence-sitter*. All of these were used when the speaker addressed the students directly, mostly to ask their opinion, e.g. *(if you're not sure you can) sit on the fence*; *who's sitting on the fence?* One Social Sciences lecturer also used *at/in the back of your mind* 7 times in the same lecture. By contrast, 13 lecturers used no idioms. Of these, 5 were from Life and Medical Sciences and 8 were from Physical Sciences.

There were many variations on established idioms in BASE. Some varied only in regard to prepositions (e.g. *at/in the back of your mind*). Others were more creative: *death by a thousand qualifications* (instead of *cuts*); *it made my year* (instead of *day*); *in one door out the other* (instead of *in one ear and out of the other*); *make sure that everybody's teaching from the sort of singing from the same hymnsheet* (*singing* is the usual word here); *so much egg on so many faces* (an intensified version of *egg on his/her face*); *preaching to the unconverted* (instead of *preaching to the converted*); *the worst of both worlds* (instead of *the best of both worlds*); and the doubly painful *shooting yourself in both feet* (instead of *shooting yourself in the foot*). Such variations add colour to lecture language, but they also make it harder for L2 students to follow the lecture or to find the idiom in a dictionary, especially if they are unfamiliar with the canonical form.

Some lecturers marked the fact that they were using figurative language by including the phrase *as it were* (e.g. sslct009 *you're sort of jumping the gun as it were*). The phrase *as it were* appeared a total of 168 times in BASE, of which 7 uses were with idioms. It did not appear at all in OCAE. However, in writing it is easier to mark phrases like this by using inverted commas, as in the OCAE example *liver biopsy-this is the 'gold standard' for the diagnosis of cirrhosis*. A similar expression, *as they say*, appeared 29 times in OCAE (0.34 pmw), 6 of which were linked to idioms (e.g. *Forewarned is forearmed, as they say*); none of the 3 BASE occurrences was linked to an idiom. Similarly, *the man/woman in the street* appeared in two BASE texts with adjectives designed to explain the meaning of the idiom: *your/the average man in the street* and *the common man in the street*. This second lecturer, from the Politics department, also used the variation *the woman in the street*. It would be interesting to see if this variation continues in more modern corpora, in light of the attempt to use more gender neutral language. Life Sciences also provided a more up-to-date version of the idiom, with 3 instances of *joe public*.

While many studies have used intuition to guess which idioms might be most useful to learn, intuition was not supported by frequency in the BASE corpus. Simpson and Mendis (2003, p. 435) suggest *in a nutshell*, for example, as a "particularly useful idiom", even though it appears only 3 times (in 2 transcripts) in MICASE. My own intuition also suggested that *in a nutshell* would appear frequently in BASE, as I had heard it used often by colleagues. It appeared only twice (1.22 pmw) in BASE, however, and was used 99 times (1.17 tokens pmw) in OCAE. These results are surprising, in that an idiom that might be supposed to be used more in speech actually occurs with similar frequency in writing. In fact, although the use of idioms in BASE was generally greater than that in OCAE, in line with Biber's (2006) claim that there is a strong difference between spoken and written university registers, many of the BASE idioms did appear in OCAE, albeit less frequently, indicating that it would be helpful for EAP students to learn the more frequent idioms for decoding purposes.

One aim of this article was to identify frequently used idioms that L2 academics could use with confidence in their own journal articles. Because frequency, as noted above, is a relative concept, it is suggested that L2 researchers familiarise themselves with the list in Table 2 and Appendix B and note where these occur in articles from their own discipline. *Gold standard*, for example, was used only in the Life Sciences in BASE and appeared 15 times in medical articles on the first page of OCAE, so academics writing for medical journals may find that the idiom is used frequently in their discipline. Once familiar with the idioms in Table 2 and Appendix B, L2 academics can then create their own lists of idioms that appear regularly in their subject fields. Teachers of L2 students, meanwhile, could use the lists in Table 2 and Appendix B to provide their students with material useful for decoding written texts, in line with the need for more materials concentrating on phraseology (Kremmel et al., 2017).

7. Conclusion

This study has identified 1393 tokens of idiom use in the BASE corpus, representing a frequency of 835 idioms pmw. The total number of individual idioms in BASE was 545, or 327 pmw. This is higher than the 260 pmw predicted by Simpson and Mendis (2003). Both this study and that by Simpson and Mendis (2003), however, confirm that idioms are used in spoken academic English and, of even greater interest, this study confirms that they are also used in written academic English. Although it is difficult to establish where idioms belong on a continuum of opacity, Martinez's (2013) recommendation that more opaque expressions (such as the idioms in this study) should be given teaching priority indicates that the more frequent idioms in this study (those appearing in 4 or more texts in BASE and at least 100 times (1.18 pmw) in OCAE) are well worth learning.

Information on the register and language variety in which idioms are found is also paramount if learners are to use them in the right situations (Liu, 2008). Future research could therefore use the BASE idioms to compare the discourse of different fields and the different types of textual organisation they present. For example, *on the other hand* clearly points to an argument presentation with contrasting points, and is used almost exclusively in BASE in Arts and Humanities texts. Given that the major limitation of this study was the lack of resources to examine the range of OCAE texts containing BASE idioms,

future studies might also address the range of the idioms in this study in OCAE and other academic writing corpora, to draw further comparisons between written and spoken registers.

Further research could also examine how many of these idioms are known and used by L2 learners and academics, and identify any factors that might prevent their comprehension and usage. For example, learners may hesitate to use an idiom because "it is as if one is claiming a cultural membership and identity one has no right to or does not wish to lay claim to" (O'Keeffe, McCarthy, & Carter, 2010, p. 223). Conversely, due to the very fact that use of an idiom marks the speaker as a "cultural insider" (Lee, 2007, p. 473), incorrect use by an L2 speaker may cause cultural dissonance for an L1 hearer. Research is therefore also needed on the attitudes of L1 speakers to L2 speakers' and writers' use of such expressions.

While future studies may shed light on the intricacies of the idioms identified here, there are two immediately practical applications of this study. First, L2 writers, students and teachers can confidently use the more frequent idioms in their own academic work, especially with regard to British English. This information is particularly useful for L2 academics, who can use these idioms to enrich their own lectures, spoken presentations and research writing. Secondly, teachers could encourage students to learn the slightly less frequent idioms for decoding purposes, using corpus examples of the idioms in sentences from academic speech and writing to provide context and enhance their learnability.

The bottom line, then, is that idioms do appear in academic discourse, and making L2 learners and researchers aware of the idioms most frequently used in this way would be, as Martinez and Murphy recommend (2011), a good use of teaching and learning time.

Acknowledgements

The recordings and transcriptions used in this study come from the British Academic Spoken English (BASE) corpus (http://www2.warwick.ac.uk/fac/soc/celte/research/base/). The corpus was developed at the Universities of Warwick and Reading under the directorship of Hilary Nesi (Warwick) and Paul Thompson (Reading). Corpus development was assisted by funding from the Universities of Warwick and Reading, BALEAP, EURALEX, the British Academy and the Arts and Humanities Research Council.

I would like to thank Russian phraseologist Dmitrij Dobrovol'skij for his patient checking of all the idioms in this study. My thanks also go to the anonymous reviewers who provided such helpful feedback on earlier drafts of the paper.

Finally, I am grateful to the School of Education and the Arts Faculty at the University of Adelaide, who each granted $2000 towards research costs for this project.

Appendix A. Idioms appearing in at least 2 BASE texts, listed by frequency in BASE

Idiom	BASE frequency per million words	BASE token count	Number of BASE texts in which the idiom occurs	Number of BASE faculties in which the idiom occurs	OCAE frequency per million words	OCAE token count
on the other hand	64.11	107	30	4	88.12	7439
bear in mind	46.73	78	42	4	10.17	859
on the one hand	37.74	63	30	4	31.87	2691
the balance of power	10.78	18	8	2	6.20	523
at the end of the day	14.98	25	13	4	1.36	115
on the other [hand]	10.19	17	12	3	20.34	1717
the bottom line	8.39	14	8	4	2.50	211
take on board	7.79	13	13	4	0.00	0
by and large	7.19	12	12	4	0.04	3
a step further/back	7.19	12	12	4	6.21	524
take for granted	7.19	12	9	3	0.01	1
in the hands of	6.59	11	10	3	12.54	1059
along the lines of	6.59	11	9	3	9.24	780
in its own right	5.99	10	10	4	9.27	783
across the board	5.99	10	7	3	1.21	102
at the back of one's mind	5.99	10	4	2	0.30	25
sit on the fence	5.99	10	2	2	0.12	10
in the long run	5.39	9	9	4	7.84	662
bad news	5.39	9	8	4	2.75	232
driving force	5.39	9	8	4	6.21	524
on the face of it	5.39	9	8	4	2.57	217
in (the) light of	5.39	9	8	3	34.99	2954
come into play	5.39	9	8	3	4.26	360
gold standard	5.39	9	6	1	6.66	562
what on earth	4.79	8	7	3	0.17	14
go without saying	4.79	8	6	2	1.85	156

(*continued on next page*)

(*continued*)

Idiom	BASE frequency per million words	BASE token count	Number of BASE texts in which the idiom occurs	Number of BASE faculties in which the idiom occurs	OCAE frequency per million words	OCAE token count
trial and error	4.79	8	5	3	1.85	156
down the line	4.19	7	7	3	0.08	7
over the top	4.19	7	6	3	0.05	4
state of the art	4.19	7	4	2	1.36	115
the man/woman in the street	4.19	7	3	3	0.28	24
stepping stone	4.19	7	2	1	0.66	56
from scratch	3.59	6	6	3	1.86	157
bridge the gap	3.59	6	6	3	1.85	156
the big picture	3.59	6	6	3	1.30	110
in the early days	3.59	6	5	4	2.33	197
get one's head (a) round	3.59	6	5	4	0.02	2
go hand in hand with	3.59	6	5	3	2.68	226
keep an eye on	3.59	6	5	3	0.78	66
hang on a minute	3.59	6	5	2	0.04	3
on the spot	3.59	6	4	4	0.84	71
get to grips with	3.59	6	4	3	1.78	150
go through the roof	3.59	6	4	2	0.06	5
full circle	3.00	5	5	3	0.54	46
that's another story	3.00	5	5	3	0.27	23
how on earth	3.00	5	5	3	0.08	7
cast one's mind back	3.00	5	5	3	0.04	3
last resort	3.00	5	5	2	4.05	342
the other side of the coin	3.00	5	5	2	0.54	46
ring a bell	3.00	5	5	1	0.15	13
good old days	3.00	5	4	3	0.27	23
grey area	3.00	5	4	2	0.60	51
out of the blue	3.00	5	3	3	0.05	4
golden age	3.00	5	3	2	2.96	250
touchy-feely	3.00	5	3	1	0.05	4
in the short run	3.00	5	2	2	4.19	354
spring to mind	2.40	4	4	3	0.31	27
on the right track	2.40	4	4	3	0.28	24
have a stab at	2.40	4	4	3	0.25	21
get the picture	2.40	4	4	3	0.08	7
the high point	2.40	4	4	2	1.52	128
it's early days	2.40	4	4	2	0.01	1
the whole story	2.40	4	3	3	1.41	119
do the job	2.40	4	3	3	0.63	53
move the goalposts	2.40	4	3	3	0.08	7
behind the scenes	2.40	4	3	2	1.36	115
in the pipeline	2.40	4	3	2	0.31	26
on the back burner	2.40	4	3	2	0.11	9
bog standard	2.40	4	3	2	0.02	2
out of one's hands	2.40	4	3	2	0.00	0
call the cavalry	2.40	4	3	2	0.00	0
beg the question	2.40	4	2	2	2.18	184
get something straight	2.40	4	2	2	0.05	4
play ball	2.40	4	2	1	0.05	4
boil down to	1.80	3	3	3	1.13	95
in store	1.80	3	3	3	0.34	29
make up one's own mind	1.80	3	3	3	0.20	17
have up one's sleeve	1.80	3	3	3	0.12	10
go down that route	1.80	3	3	3	0.09	8
get one's act together	1.80	3	3	3	0.08	7
on one hand	1.80	3	3	2	2.58	218
overall picture	1.80	3	3	2	1.14	96
golden rule	1.80	3	3	2	1.05	89
have a life of its own	1.80	3	3	2	0.89	75
turn something on its head	1.80	3	3	2	0.76	64
fall into place	1.80	3	3	2	0.30	25
so far so good	1.80	3	3	2	0.23	19

(*continued*)

Idiom	BASE frequency per million words	BASE token count	Number of BASE texts in which the idiom occurs	Number of BASE faculties in which the idiom occurs	OCAE frequency per million words	OCAE token count
on one's hands	1.80	3	3	2	0.11	9
in the same boat	1.80	3	3	2	0.09	8
take home message	1.80	3	3	1	0.11	9
joe public	1.80	3	3	1	0.04	3
kicking and screaming	1.80	3	3	1	0.02	2
rule of thumb	1.20	2	2	2	2.98	252
hot spots	1.80	3	2	2	0.94	79
ring true	1.80	3	2	2	0.30	25
put your finger on	1.80	3	2	2	0.24	20
get a handle on	1.80	3	2	2	0.21	18
shut up shop	1.80	3	2	2	0.01	1
set something in stone	1.80	3	2	1	0.12	10
swings and roundabouts	1.80	3	2	1	0.02	2
go in one ear and out the other	1.80	3	2	1	0.00	0
in a nutshell	1.20	2	2	2	1.17	99
fall foul of	1.20	2	2	2	1.15	97
set the scene	1.20	2	2	2	1.15	97
have the upper hand	1.20	2	2	2	0.88	74
on the side	1.20	2	2	2	0.84	71
make up one's mind	1.20	2	2	2	0.76	64
fly in the face of	1.20	2	2	2	0.71	60
get carried away	1.20	2	2	2	0.65	55
moot point	1.20	2	2	2	0.47	40
someone's bread and butter	1.20	2	2	2	0.38	32
stand to reason	1.20	2	2	2	0.38	32
devil's advocate	1.20	2	2	2	0.37	31
get one's message across	1.20	2	2	2	0.30	25
deliver the goods	1.20	2	2	2	0.23	19
a bad press	1.20	2	2	2	0.19	16
the powers that be	1.20	2	2	2	0.18	15
set foot in	1.20	2	2	2	0.14	12
happily ever after	1.20	2	2	2	0.11	9
shift gears	1.20	2	2	2	0.11	9
get down to the nitty gritty	1.20	2	2	2	0.09	8
in one's sights	1.20	2	2	2	0.08	7
brain power	1.20	2	2	2	0.07	6
not to mince one's words	1.20	2	2	2	0.06	5
throw somebody in at the deep end	1.20	2	2	2	0.06	5
cover one's bases	1.20	2	2	2	0.06	5
weird and wonderful	1.20	2	2	2	0.06	5
cast an eye over	1.20	2	2	2	0.06	5
above one's station	1.20	2	2	2	0.05	4
have a go	1.20	2	2	2	0.02	2
in the same ballpark	1.20	2	2	2	0.02	2
pat on the back	1.20	2	2	2	0.02	2
sit on one's hands	1.20	2	2	2	0.02	2
throw up one's hands	1.20	2	2	2	0.02	2
watch this space	1.20	2	2	2	0.02	2
go down the road of	1.20	2	2	2	0.01	1
this that and the other	1.20	2	2	2	0.01	1
get cracking	1.20	2	2	2	0.01	1
give someone a shout	1.20	2	2	2	0.01	1
have a crack at	1.20	2	2	2	0.01	1
not to put too fine a point on it	1.20	2	2	2	0.01	1
give the game away	1.20	2	2	2	0.01	1
	1.20	2	2	2	0.00	0

(*continued on next page*)

(*continued*)

Idiom	BASE frequency per million words	BASE token count	Number of BASE texts in which the idiom occurs	Number of BASE faculties in which the idiom occurs	OCAE frequency per million words	OCAE token count
beat/get the hell out of something						
get a move on	1.20	2	2	2	0.00	0
get one's thoughts together	1.20	2	2	2	0.00	0
hand on heart	1.20	2	2	2	0.00	0
quote somebody on something	1.20	2	2	2	0.00	0
put one's head above the parapet	1.20	2	2	2	0.00	0
a fair share	1.20	2	2	1	1.55	131
to say the least	1.20	2	2	1	1.10	93
grass roots movement	1.20	2	2	1	0.65	55
pick and choose	1.20	2	2	1	0.47	40
sow seeds of thought	1.20	2	2	1	0.37	31
at loggerheads	1.20	2	2	1	0.24	20
drag one's feet	1.20	2	2	1	0.21	18
in the driving seat	1.20	2	2	1	0.17	14
go back to square one	1.20	2	2	1	0.08	7
set in tablets of stone	1.20	2	2	1	0.07	6
dear to one's heart	1.20	2	2	1	0.06	5
off the top of one's head	1.20	2	2	1	0.05	4
end of story	1.20	2	2	1	0.05	4
in a rut	1.20	2	2	1	0.04	3
tick the boxes	1.20	2	2	1	0.04	3
round robin	1.20	2	2	1	0.02	2
dig one's heels in	1.20	2	2	1	0.02	2
stretch one's legs	1.20	2	2	1	0.02	2
get someone on board	1.20	2	2	1	0.00	0
on that note	1.20	2	2	1	0.00	0
tail end Charlie	1.20	2	2	1	0.00	0
jump up and down	1.20	2	2	1	0.00	0

Appendix B. BASE idioms with a frequency over 1.18 pmw (100 tokens) in OCAE, listed by frequency in OCAE

Idiom	OCAE token count	OCAE frequency per million words	BASE frequency per million words	BASE token count	Number of BASE texts in which the idiom occurs	Number of BASE faculties in which the idiom occurs
on the other hand	7439	88.12	64.11	107	30	4
in the light of	2954	34.99	5.39	9	8	3
on the one hand	2691	31.87	37.74	63	30	4
on the other [hand]	1717	20.34	10.19	17	12	3
in the hands of	1059	12.54	6.59	11	11	3
bear in mind	859	10.17	46.73	78	42	4
in its own right	783	9.27	5.99	10	10	4
along the lines of	780	9.24	6.59	11	9	3
in the long run	662	7.84	5.39	9	9	4
gold standard	562	6.66	5.39	9	6	1
driving force	524	6.21	5.39	9	8	4
balance of power	523	6.19	10.78	18	8	2
(take) a step back/further	402	4.76	7.19	12	12	4
come into play	360	4.26	5.39	9	8	3
in the short run	354	4.19	3.00	5	2	2
last resort	342	4.05	3.00	5	5	2

(*continued*)

Idiom	OCAE token count	OCAE frequency per million words	BASE frequency per million words	BASE token count	Number of BASE texts in which the idiom occurs	Number of BASE faculties in which the idiom occurs
rule of thumb	252	2.98	1.20	2	2	2
golden age	250	2.96	3.00	5	3	2
bad news	232	2.75	5.39	9	8	4
go hand in hand with	226	2.68	3.59	6	5	3
on one hand	218	2.58	1.80	3	3	2
on the face of it	217	2.57	5.39	9	8	4
the bottom line	211	2.50	8.39	14	8	4
in the early days	197	2.33	3.59	6	5	4
beg the question	184	2.18	2.40	4	2	2
bridge the gap	181	2.14	3.59	6	6	3
from scratch	157	1.86	3.59	6	6	3
trial and error	156	1.85	4.79	8	5	3
get to grips with	150	1.78	3.59	6	4	3
the good life	148	1.75	0.60	1	1	1
track record	146	1.73	0.60	1	1	1
pros and cons	145	1.72	0.60	1	1	1
raison d'ëtre	132	1.56	1.20	2	1	1
come to light	132	1.56	1.20	2	1	1
one's fair share	131	1.55	1.20	2	2	1
the high point	128	1.52	2.40	4	4	2
the whole story	119	1.41	2.40	4	3	3
win win	116	1.37	1.20	2	1	1
at the end of the day	115	1.36	14.98	25	13	4
behind the scenes	115	1.36	2.40	4	3	2
state of the art	115	1.36	4.19	7	4	2
the big(ger) picture	110	1.30	3.59	6	6	3
across the board	102	1.21	5.99	10	7	3

References

Ackermann, K., & Chen, Y.-H. (2013). Developing the academic collocation list (ACL) – A corpus-driven and expert-judged approach. *Journal of English for Academic Purposes, 12*(4), 235–247. https://doi.org/10.1016/j.jeap.2013.08.002.

Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Biber, D., & Conrad, S. (1999). Lexical bundles in conversation and academic prose. In H. Hasselgard, & S. Oksfjell (Eds.), *Out of corpora: Studies in honor of Stig Johansson* (pp. 181–190). Amsterdam: Rodopi.

de Chazal, E., & Moore, J. (2013). *Oxford EAP: A course in English for academic purposes. Advanced/C1*. Oxford: Oxford University Press.

Cowie, A. P. (1998). Phraseological dictionaries: Some east-west comparisons. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis, and applications* (pp. 209–228). Oxford: Clarendon Press.

Coxhead, A. (2000). A new academic word list. *Tesol Quarterly, 34*(2), 213–238. https://doi.org/10.2307/3587951.

Dobrovol'skij, D., & Piirainen, E. (2005). *Figurative language: Cross-cultural and cross-linguistic perspectives*. Amsterdam: Elsevier Ltd.

Durrant, P. (2009). Investigating the viability of a collocation list for students of English for academic purposes. *English for Specific Purposes, 28*(3), 157–169. https://doi.org/10.1016/j.esp.2009.02.002.

Fernando, C. (1996). *Idioms and idiomaticity*. Oxford: OUP.

Grant, L. E. (2007). In a manner of speaking: Assessing frequent spoken figurative idioms to assist ESL/EFL teachers. *System, 35*(2), 169–181. https://doi.org/10.1016/j.system.2006.05.004.

Grant, L. E., & Bauer, L. (2004). Criteria for re-defining idioms: Are we barking up the wrong tree? *Applied Linguistics, 25*(1), 38–61.

Groom, N. (2005). Pattern and meaning across genres and disciplines: An exploratory study. *Journal of English for Academic Purposes, 4*(3), 257–277. https://doi.org/10.1016/j.jeap.2005.03.002.

Howarth, P. A. (1996). *Phraseology in English academic writing: Some implications for language learning and dictionary making*. Tubingen: Max Niemeyer Verlag.

Howarth, P. A. (1998). The phraseology of learners' academic writing. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis, and applications* (pp. 161–186). Oxford: Clarendon Press.

Hyland, K., & Tse, P. (2007). Is there an "academic vocabulary"? *Tesol Quarterly, 41*(2), 235–253. https://doi.org/10.2307/40264352.

Kremmel, B., Brunfaut, T., & Alderson, J. C. (2017). Exploring the role of phraseological knowledge in foreign language reading. *Applied Linguistics, 38*(6), 848–870. https://doi.org/10.1093/applin/amv070.

Lee, P. (2007). Formulaic language in cultural perspective. In P. Skandera (Ed.), *Phraseology and culture in English* (pp. 471–496). The Hague: Mouton de Gruyter.

Lexical Computing CZ s.r.o. *Sketch engine*. n.d, Retrieved 18 June 2018 from https://www.sketchengine.eu/user-guide/glossary/.

Li, J., & Schmitt, N. (2009). The acquisition of lexical phrases in academic writing: A longitudinal case study. *Journal of Second Language Writing, 18*(2), 85–102. https://doi.org/10.1016/j.jslw.2009.02.001.

Liu, D. (2003). The most frequently used spoken American English idioms: A corpus analysis and its implications. *Tesol Quarterly, 37*(4), 671–700.

Liu, D. (2008). *Idioms: Description, comprehension, acquisition and pedagogy*. New York: Taylor and Francis.
Liu, D. (2012). The most frequently-used multi-word constructions in academic written English: A multi-corpus study. *English for Specific Purposes, 31*(1), 25–35. https://doi.org/10.1016/j.esp.2011.07.002.
Martinez, R. (2013). A framework for the inclusion of multi-word expressions in ELT. *ELT Journal, 67*(2), 184–198. https://doi.org/10.1093/elt/ccs100.
Martinez, R., & Murphy, V. A. (2011). Effect of frequency and idiomaticity on second language reading comprehension. *Tesol Quarterly, 45*(2), 267–290. https://doi.org/10.5054/tq.2011.247708.
Martinez, R., & Schmitt, N. (2012). A phrasal expressions list. *Applied Linguistics, 33*(3), 299–320. https://doi.org/10.1093/applin/ams010.
Miller J., Bringing it all together: A table of terms for multiword expressions, In: Szerszunowicz J. (Ed.), Research on phraseology across continents (Vol. 4), forthcoming, University of Bialystok Publishing House; Bialystok, Poland.
Moon, R. (1998). *Fixed expressions and idioms in English*. Oxford: Clarendon Press.
Nesi, H. (2015). ESP corpus construction: A plea for a needs-driven approach. *ASp la revue du GERAS, 68*. https://doi.org/10.4000/asp.4682.
Nesi, H., & Thompson, P. (2006). *BASE manual*. Retrieved 8 February 2019 from https://warwick.ac.uk/fac/soc/al/research/collections/base/history.
O'Dell, F., & McCarthy, M. (2010). *English idioms in use: Advanced*. Cambridge: Cambridge University Press.
O'Keeffe, A., McCarthy, M., & Carter, R. (2010). Idioms in everyday use and in language teaching. In G. Cook, & S. North (Eds.), *Applied linguistics in action: A reader* (pp. 212–230). New York: Routledge.
Park, J., & Chon, Y. V. (2018). EFL learners' knowledge of high-frequency words in the comprehension of idioms: A boost or a burden? *RELC Journal, 0*(0), 1–16. https://doi.org/10.1177/0033688217748024.
Shin, Y. K. (2019). Do native writers always have a head start over nonnative writers? The use of lexical bundles in college students' essays. *Journal of English for Academic Purposes, 40*, 1–14. https://doi.org/10.1016/j.jeap.2019.04.004.
Simpson, R., & Mendis, D. (2003). A corpus-based study of idioms in academic speech. *Tesol Quarterly, 37*(3), 419–441.
Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics, 31*(4), 487–512. https://doi.org/10.1093/applin/amp058.

Julia Miller is a senior lecturer in the School of Education, The University of Adelaide, Australia. Her main research interests are phraseology, lexicography and language learning and teaching. She is the creator of the free English for Uni website (www.adelaide.edu.au/english-for-uni), an innovative resource that aims to make English language learning fun.