THE JOURNAL OF ASIA TEFL
Vol. 13, No. 4, Winter 2016, 381-390
http://dx.doi.org/10.18823/asiatefl.2016.13.4.10.381

The Journal of Asia TEFL

http://journal.asiatefl.org/

e-ISSN 2466-1511

Opening Pandora's Box: A Corpus-Based Study of Idioms in ELT Materials

Yu-Hua Chen
University of Nottingham Ningbo, China

Longyuan Wang
University of Nottingham Ningbo, China

Introduction

Idiomatic expressions have increasingly received attention in ELT over the past few decades as they are an indispensable part of our daily communication. Depending on the degree of literalness, Fernando (1996) divides idioms into three categories: pure (non-literal), semi-literal and literal (as illustrated in TABLE 1). As pure idioms, often linguistic metaphors, tend to be understood figuratively, they are considered 'the most ambiguous multi-word expression' within the three classifications (Hummer & Stathi, 2006). Characterized by the feature of semantic opacity, pure idioms often cause great difficulty for L2 learners and are therefore the focus of the current study.

TABLE 1
Three Categories of Idioms (Fernando, 1996, p. 32)

Category	Examples
Pure	*kick the bucket, pull someone's leg, make off with*
Semi-literal	*fat chance, use something as a stepping stone, go through*
Literal	*according to, in sum, throw away*

Recent research reports that ELT materials often rely on subjective judgement and made-up sentences for the selection and introduction of idiomatic expressions, rather than real-life language data (e.g. Koprowski, 2005; Liu, 2003). As a result, these materials tend to include low-frequency lexical items that are of little pedagogical value and sometimes even fail to include proper explanations, hence negatively impacting the teaching and learning of idioms. This may be particularly true in the context of EFL, where English is taught and learned in non-English speaking countries, such as China. This issue can be addressed through corpus analysis, because corpora with large quantities of naturally occurring language data can provide a rigorous and systematic way for the identification of idioms (Biber, Johansson, Leech, Conrad & Finegan, 1999). Very little research has, however, been done comparing idioms introduced in online ELT content and that of large native corpora. This paper intends to fill this gap. To evaluate the selection and introduction of pure idioms in ELT materials regarding frequency and context of use, the current study examines a list of pure idioms

introduced by a popular ELT website based in China and compares each of the idioms against large native corpora, including the COCA (Corpus of Contemporary American English) and the BNC (British National Corpus). Through this comparative approach, the research aims to answer the following questions:

1. What are the frequencies of the chosen ELT idioms used in the reference corpora?
2. What are the contexts of use of these ELT idioms in terms of English variety and genre?
3. Based on the above findings, what are the pedagogical implications?

Literature Review

Pure Idioms

Idioms are one type of conventionalized multi-word expression, although their definition varies according to researchers and contexts (Grant, 2005). As mentioned, the present study narrows down the scope of investigation to pure idioms only. According to Simpson and Mendis (2003), pure idioms are characterised by three criteria: fixedness (i.e. fixed in a specific lexical form), institutionalization (i.e. conventionalized usage) and semantic opacity (i.e. interpretation independent of their constituent parts). The identification of idioms is, however, not always straightforward. First, despite a higher degree of fixedness, idioms are likely to occur with some form of variation, e.g. pronoun change or verb inflection (Grant, 2003, p. 116). Some idioms may occasionally undergo lexical variation or institutionalized transformation, including permutation, deletion or the addition of some elements, which makes the identification of idioms difficult. Secondly, the two major varieties of English, namely American English (AmE) and British English (BrE), may affect choices of idioms as a result of different cultures. Some idioms have been found to occur more frequently in British English than in American English, and vice versa (Tanasić, 2014). Thirdly, some idioms may have both literal and non-literal meanings. For example, '*hot potato*' can carry either literal or idiomatic meaning in different contexts. The former refers to the temperature of a potato while the latter a controversial issue or difficult situation. The context, therefore, needs to be considered when examining polysemous idioms.

Corpus Approaches in ELT

Corpus approaches have been applied in the analysis of idioms for pedagogical purposes since the 1990s. For example, Moon (1998) analysed English idioms and fixed expressions in the Oxford Hector Pilot Corpus (OHPC) and Bank of English (BoE) in terms of their frequencies, forms and functions. Biber et al. (1999) conducted a study of the Longman Grammar of Spoken and Written English Corpus (LGSWE) and created a short list of the most frequently-used idioms, and they argue that corpus analysis based on naturally occurring language data is especially helpful and productive for ELT through the evaluation of frequencies of lexical use. Furthermore, Liu (2003), Simpson and Mendis (2003) and Grant (2005) also used corpora to explore idiom selection, meaning, context and discourse function, thus contributing greatly to ELT. As Alavi and Rajabpoor (2015) criticize, however, few corpus-based studies on idioms consider real language learning contexts. This paper, therefore, aims to look into the selection and introduction of pure idioms in online ELT content and compare them to large corpora.

Methodology

Idioms Introduced on an ELT Website

Idioms from a popular Chinese ELT website called 'Shibo' were collected (n=100). Established in 1999, Shibo (www.360abc.com) is a not-for-profit website based in China dedicated to free English learning with many online resources. The idioms come from the 'Learning Idioms for Fun' section on Shibo, and

descriptions with meanings in Chinese and three to five illustrative sentences for each idiom are usually provided online. Among 100 idioms, 17 idioms are excluded from this investigation because they do not meet the criteria discussed earlier, mostly semantic opacity. The remainder comprises a list of 83 pure idioms selected for investigation.

Reference Corpora

Two reference corpora are chosen for this study: the BNC with almost 100 million words and the COCA with over 500 million words, both available from Professor Mark Davies's seminal work on Brigham Young University's website (BYU http://corpus.byu.edu). The reason for using two reference corpora is to ensure the representativeness and reliability of the comparison made against pure idioms extracted from Shibo. On the one hand, the BNC and the COCA are the only two large and well-balanced corpora which are freely available to the public. On the other hand, the data from these two corpora represent British English and American English, respectively, the two primary English varieties, which may provide valuable insights into the differences in idiom use, if any, between BrE and AmE.

Procedure

As discussed, pure idioms are characterised by semantic ambiguity and, in some cases, polysemy. In the current study, concordance lines were therefore manually checked in the case of possible polysemy, and only idiomatic meanings were considered. In addition, as idioms may have variations in form, for example pronoun or inflection change, various possible forms of the same idioms were considered when searching BYU's online interface. For example, there are two variable slots for the idiom '*bite someone's head off*': one is the verb '*bite*' while the other refers to the possessive case '*someone's*'. Following the search syntax adopted by BYU's online corpora, this idiom is then represented as '[*bite*] * *head off*', with [*bite*] indicating various lemmas of the verb '*bite*' (i.e. '*bites*', '*biting*', '*bit*', '*bitten*') and the possessive case '*someone's*' being replaced by the symbol * denoting an unknown word including '*my*', '*his*', '*their*' etc.[1]

To compare the frequency and distribution of pure idioms in Shibo and two reference corpora, the following steps were taken:

a. A list of chosen pure idioms from Shibo was compiled in a spreadsheet, and possible variations were recorded in the form of search syntax adopted by BYU's corpus interface.
b. Each of the idioms was then searched for in BYU's online BNC and COCA following the defined search syntax, and concordances of each pure idiom were reviewed to eliminate instances of literal meaning. Most pure idioms do not occur more than 100 times in the BNC or COCA, but if they do, only the first 100 random concordance lines were checked manually and then standardized frequencies were calculated and recorded according to the proportion of idiomatic sense.
c. The genres with the highest frequencies where pure idioms occur in the reference corpora were recorded. The BNC and the COCA are not distinguished here as the distributional pattern is fairly consistent between these two reference corpora. In the case of a raw frequency lower than ten, this was left blank as the data were deemed too small to yield meaningful patterns.
d. An online log-likelihood calculator (http://ucrel.lancs.ac.uk/llwizard.html) was used to see if frequency difference between the BNC and the COCA reached statistical significance.
e. The genres with the highest frequencies were identified and recorded.

[1] Note that the corpus architecture and interface on BYU's website has undergone major changes since the present study was conducted. For the most updated search syntax, see http://corpus.byu.edu.

Results & Findings

TABLE 2 is a complete list of the target pure idioms with the above quantitative information sorted by the average standardized frequency in BNC and COCA, with the most frequent idioms at the top and the least frequent ones at the bottom.

TABLE 2
Pure Idioms from Shibo with Information of Frequency and Genre Extracted from the BNC and the COCA

No	Pure Idioms from Shibo	Variation in BNC & COCA (in the form of search syntax)	Avg. Std Freq. in BNC & COCA (per m.)	Std Freq in BNC (per l.)	Std Freq in COCA (per m.)	Log-likeli hood (BNC vs COCA)	Genre with Highest Freq.	Genre with the Second Highest Freq.
1	*break the ice*	*[break] the ice*	1.17	2.23	0.11	517.44*	Magazine	Fiction
2	*red tape*	-	1.10	1.06	1.14	0.47	Newspaper	Spoken
3	*the cutting edge*	*cutting edge*	0.84	0.8	0.89	0.76	Magazine	Newspaper
4	*a free hand*	-	0.66	0.9	0.42	32.60*	Newspaper	Magazine
5	*black sheep*	-	0.59	0.88	0.29	56.89*	Fiction	Magazine
6	*drag one's feet*	*[drag] * feet*	0.58	0.56	0.60	0.20	Fiction	Newspaper
7	*read between the lines*	*[read] between the lines*	0.56	0.65	0.46	5.41	Spoken	Fiction
8	*a Pandora's box*	*Pandora's box*	0.53	0.35	0.71	18.77*	Newspaper	Spoken
9	*on thin ice*	-	0.37	0.71	0.02	181.84*	Magazine	Fiction
10	*a close call*	-	0.34	0.23	0.45	11.39	Spoken	Fiction
11	*Achilles' heel*	-	0.30	0.18	0.41	14.26*	Magazine	Newspaper
12	*let off steam*	*[let] off steam*	0.29	0.4	0.18	16.14*	Magazine	Newspaper
13	*keep a straight face*	*[keep] a straight face*	0.27	0.28	0.26	0.12	Fiction	Spoken
14	*hit and miss*	*[hit] and [miss]*	0.26	0.29	0.24	0.95	Spoken	Fiction
15	*dark horse*	-	0.25	0.16	0.34	10.50	Newspaper	Spoken
16	*let your hair down*	*[let] * hair down*	0.24	0.27	0.21	1.21	Fiction	Spoken and Magazine
17	*white elephant*	-	0.24	0.28	0.19	2.72	Newspaper	Spoken
18	*under the weather*	-	0.23	0.31	0.15	9.91	Fiction	Magazine
19	*eat one's words*	*[eat] * words*	0.20	0.25	0.15	4.04	Spoken	Newspaper
20	*blow one's top*	*[blow] * top*	0.19	0.26	0.13	7.98	Magazine	Newspaper
21	*white lie*	-	0.19	0.14	0.24	3.97	Fiction	Spoken
22	*bite sb's head off*	*[bite] * head off*	0.18	0.22	0.13	3.94	Fiction	Spoken
23	*get cold feet*	*[get] cold feet*	0.18	0.14	0.23	3.37	Spoken	Fiction
24	*go to the dogs*	*[go] to the dogs*	0.18	0.19	0.17	0.16	Spoken	Newspaper
25	*Spring chicken*	-	0.18	0.1	0.25	10.09	Newspaper	Fiction
26	*red-letter day*	*red-letter day / red letter day*	0.15	0.19	0.10	4.70	Fiction	Newspaper
27	*straight from the horse's mouth*	*the horse's mouth*	0.15	0.19	0.10	4.99	Spoken	Fiction
28	*once in a blue moon*	-	0.14	0.16	0.12	1.09	Fiction	Newspaper
29	*paper tiger*	-	0.14	0.14	0.15	0.04	Spoken	Newspaper
30	*smell a rat*	*[smell] a rat*	0.14	0.15	0.12	0.56	Fiction	Spoken

31	*hot potato*	-	0.13	0.09	0.17	4.12	Spoken	Newspaper
32	*keep one's eye on the ball*	*eye on the ball*	0.13	0.07	0.19	8.22	Spoken	Magazine
33	*pay through the nose*	*[pay] through the nose*	0.13	0.16	0.09	3.33	Fiction	Newspaper
34	*call a spade a spade*	*[call] a spade a spade*	0.12	0.12	0.12	0.00	Fiction	Spoken
35	*green thumb*	-	0.12	0.01	0.24	34.62*	Magazine	Fiction
36	*go Dutch*	-	0.11	0.17	0.06	10.68	Magazine	Newspaper
37	*hot ticket*	-	0.11	0.02	0.20	24.38*	Newspaper	Spoken
38	*let sleeping dogs lie*	*[let] sleeping dogs lie*	0.11	0.12	0.09	0.67	Spoken	Newspaper
39	*curry favour*	*[curry] favour*	0.10	0.2	0.00	60.55*	Newspaper	Non-academic
40	*keep your fingers crossed*	*[keep] your fingers crossed*	0.10	0.12	0.08	1.56	Spoken	Magazine and Fiction
41	*knock on wood*	*[knock] on wood*	0.10	0	0.20	36.92*	Spoken	Magazine
42	*let the cat out of the bag*	*cat out of the bag*	0.10	0.13	0.07	3.06	Fiction	Spoken
43	*talk turkey*	*[talk] turkey*	0.10	0.06	0.14	4.54	Fiction	Spoken
44	*burn the midnight oil*	*[burn] the midnight oil*	0.09	0.1	0.08	0.22	Magazine	Fiction
45	*wet behind the ears*	-	0.09	0.11	0.06	2.09	Fiction	Magazine
46	*cost an arm and a leg*	*[cost] an arm and a leg*	0.08	0.08	0.07	0.05	Fiction	Magazine
47	*make your hair stand on end*	*[make] * hair stand on end*	0.08	0.11	0.05	3.93	Fiction	Magazine
48	*blow hot and cold*	*[blow] hot and cold*	0.07	0.12	0.03	12.46	Magazine	Spoken
49	*behind the eight ball*	-	0.05	0.02	0.09	6.79	Newspaper	Spoken
50	*Dutch courage*	-	0.05	0.09	0.01	12.90	Fiction	Spoken
51	*play cat and mouse*	*[play] cat and mouse*	0.05	0.05	0.05	0.02	Magazine	Spoken
52	*a bull in a china shop*	-	0.04	0.05	0.04	0.29		
53	*at sixes and sevens*	-	0.04	0.06	0.01	7.31	Fiction	Spoken
54	*dressed up to the nines*	*[dress] up to the nines*	0.04	0.07	0.00	23.87*		
55	*follow your nose*	-	0.04	0.06	0.03	2.36	Spoken	Magazine
56	*laugh like a drain*	*[laugh] like a drain*	0.04	0.08	0.01	15.59*	Fiction	Magazine
57	*rain cats and dogs*	*[rain] cats and dogs*	0.04	0.03	0.05	1.02	Fiction	Spoken
58	*blow one's own horn*	*[blow] * own horn*	0.03	0	0.05	9.23	Newspaper	Spoken
59	*a snake in the grass*	-	0.02	0.02	0.02	0.00	Fiction	Magazine
60	*bell the cat*	*[bell] the cat*	0.02	0.02	0.01	0.23		

61	*a dog in the manger*	-	0.01	0.01	0.00	3.41		
62	*a fat cat*	-	0.01	0	0.03	4.82	Spoken	Fiction
63	*a monkey on my back*	-	0.01	0	0.02	4.01	Spoken	Fiction
64	*between cup and lip*	-	0.01	0.01	0.00	3.41		
65	*carry coals to Newcastle*	*[carry] coals to Newcastle*	0.01	0.01	0.01	0.11		
66	*every dog has his day*	-	0.01	0.01	0.00	0.39		
67	*give a black eye*	*[give] * a black eye*	0.01	0	0.01	2.01		
68	*green hand*	-	0.01	0.01	0.01	0.11		
69	*Indian giver*	-	0.01	0	0.01	2.41		
70	*six and two three*	-	0.01	0.01	0.00	3.41		
71	*sixty-four dollar question*	-	0.01	0.01	0.00	3.41		
72	*take French leave*	*[take] French leave*	0.01	0.01	0.00	1.04		
73	*under the rose*	-	0.01	0	0.03	5.22	Fiction	Spoken
74	*cast pearls before swine*	*[cast] pearls before swine*	0.00	0	0.00	0.80		
75	*cat and dog life*	-	0.00	0	0.00	0.00		
76	*find one's feet*	*[find] * feet*	0.00	0	0.00	0.40		
77	*have irons in the fire*	*[have] irons in the fire*	0.00	0	0.00	0.80		
78	*laugh in one's sleeve*	*[laugh] in * sleeve*	0.00	0	0.00	0.00		
79	*nine days' wonder*	-	0.00	0	0.00	0.00		
80	*pull one's leg*	*[pull] * leg*	0.00	0.01	0.00	1.40		
81	*run with the hare and hunt with the hounds*	*[run] with the hare and [hunt] with the hounds*	0.00	0	0.00	0.00		
82	*show the cloven foot*	*[show] the cloven foot*	0.00	0	0.00	0.00		
83	*show the white feather*	*[show] the white feather*	0.00	0	0.01	1.20		

Note. An asterisk * indicates statistical significance ($p<0.001$).
The symbol - indicates the search string has no variation.

Frequency

It is often argued that frequency of occurrence should be one of the key parameters when choosing lexical items for teaching (Nation & Macalister, 2010), and it is, therefore, sensible to prioritise high-frequency idioms in any ELT syllabus. As can be seen in TABLE 2, the majority of pure idioms occur less than once per million words, which supports Moon's argument (1998) that idioms are generally rare, although this does not mean that idioms are not important for ELT. Meanwhile, TABLE 2 also indicates that the frequencies of pure idioms vary considerably. Some idioms are relatively frequent (e.g. '*break the ice*' and '*red tape*'), whereas some scarcely occur in the reference corpora (e.g. '*take French leave*' or '*between cup and lip*'). Twenty-eight

per cent of pure idioms occur no more than 0.01 times per million words, which means a significant proportion of the pure idioms introduced by Shibo occur less than ten times in the very large collections of texts in the BNC and COCA (around 600 million words in total). It appears that there is no systematic selection of idioms on the basis of frequency in real-life use when considering which idioms to teach, and it also suggests that the chances of L2 students coming across those low-frequency idioms are very slim. As a matter of fact, the following five idioms introduced by Shibo do not occur at all in the BNC or COCA: '*cat and dog life'*, '*laugh in one's sleeve*', '*nine day's wonder*', '*run with the hare and hunt with the hounds*' and '*show the cloven foot*'. To address this issue, it is recommended to prioritise higher-frequency pure idioms and discard those with low frequency, such as those occurring less than 0.01 times per million words. A theoretical framework proposed by Martinez (2013) provides a good starting point when considering the inclusion of multi-word expressions such as idioms in ELT materials or syllabuses.

British English and American English

The webpages on Shibo which introduce pure idioms indicate that no information in relation to context (i.e. the context that each idiom often occurs in) is provided. It should be noted that idiom choice varies between British and American linguistic communities as well as across different types of texts. As mentioned, Tanasić (2014) confirms that the BNC and the COCA reflect differences in BrE and AmE because of the sources of texts. In the current study, as can be seen in TABLE 2, some pure idioms have drastically different frequencies in the BNC and the COCA. Take '*break the ice*' for example: it occurs 223 times in the BNC but only 51 times in the much larger corpus COCA. As the BNC and the COCA are of different sizes, to assess whether the difference is statistically significant the log-likelihood (LL) test was used. Pure idioms with an LL score greater than 13.5 – the critical value for significance ($p<0.001$) – are tabulated below in TABLE 3. Pure idioms on the left are significantly more frequent in BrE whereas those on the right are significantly more frequent in AmE. ELT material writers can therefore use different corpora to highlight the distinctiveness of certain idiom use across different English varieties.

TABLE 3
Pure Idioms with Significant Differences in Frequency between the BNC and COCA Sorted by LL Scores

More frequent in the BNC	LL score	More frequent in the COCA	LL score
break the ice	517.44	*black sheep*	56.89
on thin ice	181.84	*knock on wood*	36.92
curry favour	60.55	*green thumb*	34.62
dress up to the the nines	23.87	*a free hand*	32.60
let off steam	16.14	*hot ticket*	24.38
laugh like a drain	15.59	*Pandora's box*	18.77
		Achilles' heel	14.26

Genre

According to BYU's corpus website, the BNC contains seven types of text, including spoken, fiction, magazine, newspaper, non-academic, academic and miscellaneous, while the COCA is evenly divided among spoken, fiction, magazine, newspaper and academic journal. The genres with the highest frequency where each of the idioms occurs are presented in TABLE 2, although the idioms with fewer than ten occurrences (raw frequency before standardization) are not considered. As the frequency distribution across genres suggests (Figure 1), the pure idioms investigated occur most often in the genres of fiction and spoken English (transcripts of unscripted conversations), and then magazine and newspaper. Their use across genres to some extent demonstrates that pure idioms can be used in both written and spoken texts but not in academic writing, as this genre exists for both the BNC and the COCA. For pedagogical purposes, typical sample sentences can then be extracted from representative genres that individual idioms most occur in, rather than made-up sentences which may not be authentic and representative of real-life language.

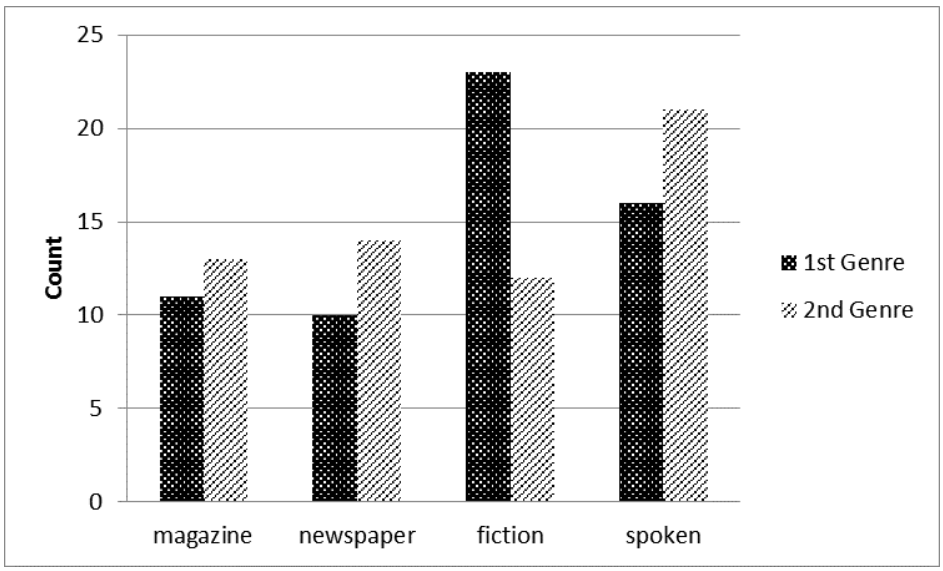

Figure 1. Distribution of idiom use in the genres with the highest frequencies.

Discussion & Conclusion

This corpus-based study has evaluated the selection and introduction of pure idioms on a Chinese ELT website by examining frequency and context in two large reference corpora. In terms of frequency, over a quarter of pure idioms introduced by Shibo are scarce, with frequency of occurrence no more than 0.01 per million tokens in the reference corpora, and such low frequency suggests that L2 learners may never come across those idioms in real-life communications. With respect to the context of idiom use in English varieties, 13 idioms occurred with statistically significant differences in frequency between British English and American English, with around half of them more frequent in BrE (e.g. '*break the ice*') and the other half more frequent in AmE (e.g. '*knock on wood'*). In addition, most of the pure idioms are frequently used in the genres of fiction, spoken English, magazine and newspaper (in order of preference), but not in academic writing. This kind of distributional information has important implications for pedagogy but is not mentioned on the Chinese ELT website.

Based on the above findings, this corpus-based investigation suggests that ELT materials should include idioms on the basis of frequency in large reference corpora, and that contextual information, such as English variety and genre, can also be used to enhance L2 acquisition of pure idioms. This does not mean that frequency should be the only criterion to be considered when introducing idioms. Some may argue that the extent to which the idioms can arouse students' interest is equally or even more important than frequency alone. For example, '*rain cats and dogs*' only occurs 0.04 per millions words on average, but teachers may be able to tell a story about this idiom which generates a visual mental image and thus trigger students' curiosity about such non-literal use of idioms. Other idioms in relation to cats and dogs such as '*fight cat and dog*' can also be introduced together to strengthen students' interest. This paper, however, contends that the information of frequency and context should always accompany the introduction of each idiom so that students are well informed and can choose to prioritise the idioms which may be of more relevance to their own needs. It should also be noted that the presentation of such information needs to be carefully considered. Take frequency for example. As we cannot expect students to understand what frequency means out of context, it may be possible to divide idioms into three frequency bands, i.e. high (frequency $\geq$0.15 per mill), mid (0.01 per mill < frequency < 0.15 per mill), and low (frequency $\leq$0.01 per mill). Although the cut-off points of such frequency bands may seem somewhat arbitrary at this stage and require further validation, they can provide a

general rule of thumb for practitioners to assess the usefulness of idioms in relation to frequency.

For future research, lexico-grammatical profiles, for example collocation or semantic prosody (Sinclair, 1996) extracted from corpus evidence, can be provided to enrich ELT instructional content on pure idioms. Martinez's framework (2013) for the inclusion of multi-word expressions can also be used to assess the selection of idioms in ELT materials.

The Authors

Dr Yu-Hua Chen (corresponding author) is interested in how corpus analyses can be used to facilitate or validate the approaches to teaching or assessing language skills. Dr Chen's research has been published in international journals including *Applied Linguistics*, *English for Academic Purposes*, and *Language Learning and Technology*.

School of English
University of Nottingham Ningbo
199 Taikang East Road
Ningbo 31500, China
Tel: +86 (0)574 8818 0000 ext. 9690
Fax: +86 (0)574 8818 0449
Email: Yu-Hua.Chen@nottingham.edu.cn

Longyuan Wang received a BA degree in English with International Business at The University of Nottingham Ningbo China in 2015. She undertook this pilot corpus project supervised by Dr Yu-Hua Chen.

School of English
University of Nottingham Ningbo
199 Taikang East Road
Ningbo 31500, China
Email**:** Shirleywangly@hotmail.com

References

Alavi, S., & Rajabpoor. A. (2015). Analysing idioms and their frequency in three advanced ILI textbooks: A corpus-based study. *English Language Teaching, 8*(1), 170–179.

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

Fernando, C. (1996). *Idioms and idiomaticity*. Oxford, UK: Oxford University Press.

Grant, L. (2003). *A corpus-based investigation of idiomatic multiword units* (Unpublished doctoral dissertation). The Victoria University of Wellington, New Zealand.

Grant, L. (2005). Frequency of core idioms in the British National Corpus (BNC). *International Journal of Corpus Linguistics*, *10*(4), 429–451.

Hummer, C., & Stathi, K. (2006). Polysemy and vagueness in idioms: A corpus-based analysis of meaning. *International Journal of Lexicography*, *19*(4), 361–377.

Koprowski, M. (2005). Investigating the usefulness of lexical phrases in contemporary coursebooks. *ELT, 59*(4), 322–332.

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

Martinez, R. (2013). A framework for the inclusion of multi-word expressions in ELT. *ELT, 67*(2), 184–198.

Moon, R. (1998). *Fixed expressions and idioms in English*. Oxford, UK: Clarendon Press.

Nation, I. S. P., & Macalister, J. (2010). *Language curriculum design*. New York & Abingdon, Oxon: Routledge.

Simpson, R., & Mendis, D. (2003). A corpus-based study of idioms in academic speech. *TESOL Quarterly, 37*(3), 419–442.

Sinclair, J. (1996). The search for units of meaning. *Textus, 9*(1), 75–106.

Tanasić, N. (2014). Arm and leg idioms in the BNC and COCA corpora: Views on the cultural differences between British and American society. *International Journal of Cognitive Research in Science, Engineering and Education (IJCRSEE)*, *2*, 77–86.

ELT & Corpus Website

Davies, M. (2004). *BYU-BNC*. (Based on the British National Corpus from Oxford University Press). Available online at http://corpus.byu.edu/bnc/. [Accessed in December 2015].

Davies, M. (2008). *The Corpus of Contemporary American English: 520 million words, 1990-present*. Available online at http://corpus.byu.edu/coca/. [Accessed in December 2015].

Shibo. (2015). *Learning idioms for fun*. Available online at http://www.360abc.com/article/list_111_6.html. [Accessed in December 2015].