Analyzing Idioms and Their Frequency in Three Advanced ILI Textbooks: A Corpus-Based Study

Sepideh Alavi[1] & Aboozar Rajabpoor[1]

[1] Department of Foreign Languages and Linguistics, Shiraz University, Iran

Correspondence: Sepideh Alavi, Department of Foreign Languages and Linguistics, Shiraz University, Iran. E-mail: salavi@rose.shirazu.ac.ir

Abstract

The present study aimed at identifying and quantifying the idioms used in three ILI *Advanced* level textbooks based on three different English corpora; MICASE, BNC and the Brown Corpus, and comparing the frequencies of the idioms across the three corpora. The first step of the study involved searching the books to find multi-word idiomatic expressions used in each. Idioms matching criteria for idiomaticity were selected and searched in the three online corpora to find their frequency of occurrence. Chi-square tests were then run to discover whether there were significant differences among the frequencies of occurrence of each idiom across each corpus. Having the number of idioms in each textbook, two other chi-square tests were then run, the first aiming at finding out if there were any significant differences among the three books in terms of idiom types and the second, to compare their tokens. The results showed that the books were different in terms of both number and type of idioms. It was also found that the idioms chosen for these *Advanced* level books did not meet necessary frequency criteria according to the literature, which could be attributed to representativeness issues of the corpora or their scope in terms of language level, genre and speaker's age.

Keywords: English idioms, online corpora, frequency

1. Introduction

Idiomatic expressions are inseparable parts of each language in both written and spoken forms, and teaching them is important in every foreign language (FL) or second language (L2) learning situation. For this reason, it seems imperative for materials developers and teachers to identify and include the most relevant idioms in their SL/FL materials and instruction. To this end, a solid definition for the concept of idiom must be provided before the proper idioms could be selected.

The word idiom has been defined by scholars in different ways. Moon (1998), for instance, uses the term in the narrow sense to refer to multi word expressions which are "not the sum of their parts" (p. 4) and whose meaning cannot be retrieved from the individual meanings of the component words. Similarly, Sporleder, Linlin, Gorinski and Koch (2010), define idioms as "multi word expressions whose meanings cannot be inferred from the meaning of their parts in a completely compositional manner" (p. 1). Simpson and Mendis (2003) pooled and summarized these definitions and identified an idiom as "a group of words that occur in a more or less fixed phrase whose overall meaning cannot be predicted by analyzing the meaning of its constituent parts." (p. 419). According to Fernando (1996), McCarthy (1998), and Moon (1998), other conditions should also be met if a multi word expression (MWE) is to qualify as an idiom; institutionalisation (the degree to which an idiom is conventionalized), fixedness (the flexibility of word sequences in an idiom), and semantic opaqueness (the unfeasibility of interpretation of the idiom based on its constituent parts).

Selection of the right idioms is important when it comes to classroom teaching and L2 materials development. To this end, many SL/FL educators act on their intuition and prior knowledge and make choices based on their personal experience, topic, key words and metaphoric themes. However, researchers such as Gardner and Davis (2007), Grant (2005), Liu (2003), Minugh (2002) and Simpson and Mendis (2003) have favored using language corpora as reliable sources for selecting idioms rather than "unprincipled and idiosyncratic" (p. 423) individual methods. They suggest finding idioms which are most frequent in the corpora and including them in ESL textbooks. They believe that the resulting selection will be objective and free from personal attitudes, tastes and opinions. In addition, students will be able to benefit more from a course including vocabulary which is more

frequently used in real life and more relevant to their needs.

On the other hand, with the bulk of material (linguistic and non-linguistic) that has to be learned by students in short periods of time, selective, efficient learning becomes a goal in itself. In other words, students might feel the need to spend their learning time on items which are more likely to be used in their future encounters with the L2, rather than less frequent and less practical ones.

For the reasons mentioned above, it seems necessary for teachers and materials writers to set frequency priorities when selecting and authoring ELT materials. According to Leech (2001), such selection seems to be a matter of common sense; however, it has been much neglected in the actual process of materials development because of the limited knowledge of course designers and the lack of expert attention to course content. This can also be true for materials of certain language institutes which are written specifically for in-school purposes.

One of the largest language schools of Iran is the Iran Language Institute (ILI). It operates in 27 provinces and 73 cities, and has over 200 branches across the country. Due to the large number of ESL students studying at this school every year, and the success and popularity the institute has gained over other language centers, it seems necessary to examine, review and if necessary, revise the textbooks in order to maximize efficient language learning. Such reviewing process could target course syllabi holistically, or deal with more discrete and detailed grammatical, semantic, or lexical aspects. The purpose of the present work was to find and examine the idioms used in three advanced level ILI books. To this end, we first identified the idioms included in these textbooks based on the definition of idioms provided by Simpson and Mendis (2003). Next, the idioms were checked in one British English corpus, the British National Corpus (BNC), and two American spoken and written corpora, the Michigan Corpus of Academic Spoken English (MICASE) and the Brown University of Standard Corpus of Present-Day American English (Brown Corpus), to find out how frequently they occurred in these corpora, and whether their inclusion in the textbooks is realistic and reasonable and corresponds to empirical evidence of actual language use.

2. Background

According to Leech (1997, cited in Garside, Leech, and McEnery (Eds.), being rich sources for materials development, corpora can bridge language teaching and learning indirectly, assuring both teachers and students that the language being used in textbooks is contemporary, useful and similar to what they are most likely to encounter in their future use of L2. As such, the activities in corpus-informed materials can focus on the most important features of language skills and produce more effective communication (McCarthy, 1998).

According to O'Keeffe, McCarthy and Carter (2007), Aijmer (2009), and Campoy, Gea-valor and Belles-Fortuno (2010), corpus-based studies can be applied in several areas of language pedagogy and classroom research. One of the particular areas of interest of corpus linguistics researchers is the use of quantitative data to obtain information about vocabulary items and how they are used in a language in the form of frozen forms such as collocations, phrasal verbs and idioms (Mindt, 1996). Such studies aim at assisting teachers to create materials that correspond to real, authentic language use, and learners to communicate more successfully using the most common words and expressions used in the target language McEnery and Xiao (2011).

Several researchers have attempted to analyze and categorize data from existing corpora and to create lists from which vocabulary items could be selected for language instruction. Gardner and Davis (2007), for example, used the BNC to identify the most frequent English phrasal verbs to be taught to EFL/ESL students. Trebits (2009) also explored the use of phrasal verbs in English language documents of the European Union to serve as a basis for the compilation of teaching materials designed to develop the necessary language skills of those who work with English language EU documents.

McCarthy (1998), used the 5-million-word CANCODE (Cambridge and Nottingham Corpus of Discourse in English) to categorize the basic spoken vocabulary into nine levels including basic parts of speech such as basic nouns, basic adjectives, basic adverbs and basic verbs for action and events, as well as modal items, de-lexical verbs, interactive words, discourse markers and generic deictics. He suggested that creating word lists from linguistic corpora can result in a "more use-centered vocabulary pedagogy at the elementary level and provide useful and usable language items even to very low level learners" (McCarthy, 1998, p. 20).

In the area of collocation studies, Kennedy (2003) used the British National Corpus (BNC) to show the nature of English collocations, how they were structured, and how they should be taught in a FL/SL situation. Later on, Shin and Nation (2008), who defined collocations as a group of unrestricted words that co-occurred, presented a list of the most frequent collocations of spoken English, again using the BNC. Walker (2011) also used two corpora to identify and cross-check useful collocations for students of business English; the Bank of English

corpus and the financial and commercial sections of the BNC. He compared the most frequent collocations obtained from each corpus and found differences between the ways words collocated in general and business English.

2.1 Research on Idioms

Until recently, corpus related idiom studies have been rather scarce because of what Ellis (1985) refers to as stronger emphasis on grammar as compared to vocabulary. However, this situation started to improve in the 1990s, resulting in exceptional works such as Moon (1998) who studied and analyzed idioms from a text-based point of view using corpus evidence on idiom frequencies, forms and functions.

Minugh (cited in Kirk, 2000) analyzed five newspaper CD-ROMs from 1995 (about 20-25 M words each) to find out how well the idioms matched with each other and with the Bank of English and how their distributions could be compared to those from the Bank of English. He found that the idioms in these newspapers matched very closely with each other and with the Bank of English in terms of frequency, and had similar distribution of idioms.

One of the other comprehensible works carried out on idiom corpus studies is Biber, Johansson, Leech, Conrad and Finegan's (1999), who analyzed the 40-million-word Longman Spoken and Written English Corpus and created a short list of the most frequently used idioms.

Although the mentioned works were unique in their approach towards studying idioms, they did not directly relate to classroom practice. According to Liu (2003), most teaching materials written on English idioms were primarily based on the teachers' or materials writers' intuition. As such, they often include rarely used idioms and sometimes even incorrect meanings. Liu (2003) addressed this problem by searching and analyzing the idioms used in Corpus of Spoken, Professional American English (Barlow, 2000), the MICASE (Simpson, et al. 2002), and Spoken American Media English (Liu, 2003). After analyzing the results, he compiled four lists of the most frequently used idioms and managed to uncover patterns of idiom use and inadequacies of the existing idiom teaching and reference materials in terms of item selection, meaning and use, and the appropriateness of the examples provided.

Simpson and Mendis (2003), also searched a corpus of 1.7 million words (MICASE) for idioms and studied their pragmatic functions and cross functions such as evaluation, description, paraphrase, emphasis, collaboration and metalanguage. They conclude that language teachers should construct classroom materials based on frequency counts and raise student awareness regarding the idioms' context of use and discourse functions. They suggested a combination of holistic and analytic approaches to teaching idioms in authentic discourse and sociopragmatic contexts to help improve their learning.

Grant (2005) also used the BNC to develop a comprehensive list idiom. The results of his corpus search, however, showed that none of the idioms identified by the analysis occurred as frequently as the most frequent 5,000 words of English. He concluded that teachers could help students recognize idiomatic expressions, using dictionaries that provide further examples of their meaning and use.

2.2 Corpus-Based Studies Put into Practice

The studies reviewed above have all focused on certain elements of language using corpus analysis to create and prescribe learning lists of vocabulary, phrasal verbs, collocations and idioms; however, very few researchers have been interested in putting the findings in a real language learning context. Among the enormous series of textbooks developed for teaching English as a foreign language, only a handful have adopted corpus-based approaches as means of selecting vocabulary and idioms. An instance of such textbooks is the well-known *English in Use* series, by Cambridge University, which cover various areas of vocabulary such as phrasal verbs, idioms and collocations which have been found to be more frequent in the English language based on a 250-million-word corpus of spoken and written English, taken from newspapers, novels and magazines, as well as more public sources.

The Touchstone series, by McCarthy (2004), is a major English course book series based on the North American English portion of the Cambridge International Corpus, claiming to have used the most frequent grammar structures and vocabulary across the corpus.

University Language by Biber (2006) is another book aiming at university registers using the T2K-SWAL corpus of 2.7 million words collected from four universities across the United States during class sessions and office hours.

Despite the small number of textbooks compiled based on corpus analysis, the method has recently gained

momentum, and it seems that publishers such as Cambridge University and Oxford University are heading towards the use of corpora as rich sources of vocabulary and/or MWEs for the materials they prepare. It is, therefore, imperative for materials developers and teachers to be aware of the importance of word frequency in the process of creating materials so that they could make better judgments and choices when selecting vocabulary for instruction. Another implication of such works for teachers is to enable them to analyze the new MWE of the books they teach so that they can enrich their instruction by adding more frequent items and MWEs in case the textbook they use needs improvement in that area.

3. The Present Study

In spite of the numerous corpus-based studies done in the field of linguistics and language teaching, few Iranian researchers have paid attention to this growing aspect of materials development. As a response to the lack of corpus-based textbooks and corpus-based studies on currently used textbooks in Iranian institutes, the present study aimed at carrying out a corpus-based research on the Iran Language Institute advanced level textbooks. Given that idiom use is considered as an indicator of fluency in a language, the focus of the present study was to find out if the idioms used in the three ILI advanced books were also frequently used in real corpora. In addition, the study intended to compare the frequencies of the idioms across three popular English corpora; the MICASE, the BNC, and the Brown Corpus and to find out if any underlying pattern governing the selection of the idioms existed. To achieve these goals, the researchers aimed at finding the answers to the following two questions:

1) How many idioms are there in each of the three ILI advanced textbooks? Is there any significant difference between the three volumes in terms of the number of idioms used?

2) How frequent are the idioms included in the three ILI textbooks in MICASE, BNC and Brown corpora?

The present study is significant as the results can provide users of ILI textbook writers, teachers and students with important information regarding the frequency of idioms included in these textbooks. Such information can help improve learning outcomes by developing materials that enhance better communicative competence (Leech 2001) and providing better range, coverage and learnability of the target language elements (van Els et al., cited in Leech, 2001).

4. Method

4.1 Textbooks

The idioms to be analyzed for frequency of use were extracted from *Advanced 1, 2 and 3* books which were planned, compiled and revised by the research and planning department of the ILI in 2007. The adult section of ILI consists of Basic, Elementary, Pre-Intermediate, Intermediate, High-Intermediate and Advanced levels, each of which contains three sub-levels and three textbooks. The three advanced textbooks analyzed here, advanced 1, 2 and 3, are taught in the last three levels of the ILI. These textbooks have a similar number of pages (pp. 129-134). All contain six chapters and two progress tests; one in the middle and one at the end of each textbook. Each chapter has four sections, each covering listening, reading, speaking and writing activities. The textbooks were reviewed thoroughly to find MWE expressions which could fit into the description of idioms.

4.2 Corpora

Since the three ILI advanced level textbooks contained both reading passages and dialogs as samples of written and spoken British and American English, corpora derived from the same language forms were needed for closer comparisons and frequency counts. Hence, the MICASE Corpus and the Brown Corpus were selected as bodies of American English speech and writing, respectively. The British National Corpus was also selected as a sample of both written and spoken British English. Other reasons for selecting these corpora were their popularity in English corpus analysis as well as their availability and ease of access and use through the internet.

4.2.1 MICASE

MICASE is a specialized corpus of contemporary American English speech recorded at the University of Michigan between 1997 and 2001 by Simpson, Briggs, Ovens, and Swales in 2002 which is freely available and searchable via the Internet. MICASE contains 197 hours of recorded speech, totaling about 1.7 million words in 152 speech events.

4.2.2 The Brown Corpus

Another corpus used is the Brown University Standard Corpus of Present-Day American English (Brown Corpus) that was compiled in the 1960s by Francis and Kucera at Brown University as a general corpus (text collection) in the field of corpus linguistics. It contains 500 samples of English-language texts, from 15 different genres, totaling roughly to one million words, compiled from works published in the United States in 1961.

4.2.3 The British National Corpus

The third corpus is the BNC formed in 1990 by the Longman British Library which started to produce a hundred million word corpus of modern British English for use in commercial and academic research. The full BNC contains about 100 million words: 90% written and 10% spoken texts. The first version of this corpus was released in 1995, including samples from different sources and genres.

4.3 Procedure

The idioms were selected based on the criteria provided by Fernando (1996). Similar to what Simpson and Mendis (2003) did, each of the expressions existing in the textbooks was tested against the three features, namely, compositeness or fixedness, institutionalization, and semantic opacity to make sure whether the expression could be considered an idiom.

Phrasal verbs were also considered as idioms because many of them are fixed in structure and non-literal or semi-literal in meaning. Verb-plus-particle or verb-plus-preposition structures that did not follow the definition of phrasal verbs were excluded from the analysis.

To determine whether a verb-plus-particle structure was a phrasal verb or not, the researchers adopted the testing method suggested by Celce-Murcia and Larsen-Freeman (1999), consisting of three criteria: the plausibility of adverb insertion between the verb and particle, the absence of literal meanings for the constituting parts, and the possibility of particle forefronting in sentences. The application of these testing principles excluded phrasal verbs such as *come in*, *go out*, *listen to*, *look at*, and *talk about*.

Based on the above criteria, 18 idioms were found in *Advanced 1*, 55 in *Advanced 2* and 42 in *Advanced 3*. These idioms were then compared by running chi-square tests using SPSS version 16. The first aimed at finding out if there were any significant differences among the three textbooks in terms of idiom type, and the second was run to compare the tokens. It should be noted that in this study, both tokens and types of idioms were taken into consideration.

Another chi-square test was run to find the frequency of occurrence of each idiom in the MICASE, BNC and Brown corpora online to discover whether there were any differences among the frequencies of occurrence of each idiom across the three corpora

Afterwards, given that the corpora were of various sizes, the researchers equalized the number of tokens as per one million words. Besides, taking the three corpora all together as one larger corpus, the researchers sought the frequency of each idiom across this corpus. Finally, based on Liu (2003) and Moon (1998), the idioms were classified into three frequency-of-use bands representing <50, 20–49 and 2–19 tokens per million words. Since there were many idioms in the textbooks which had frequencies lower than 2 per million, another band, not present in Liu's study, was added to include these infrequent idioms. It is worth mentioning that the token of each idiom in the entire corpus was the cornerstone for categorizing that idiom in a specific band.

5. Results

To answer the first research question regarding the number of idioms in the ILI advanced textbooks, a total of 116 idiom types and a total of 159 idiom tokens were found in the three textbooks. The results are presented in Table 1.

Table 1. Number of idiom types and tokens in each textbook

Textbook	Number of idiom types	Number of idiom tokens
Advanced 1	19	26
Advanced 2	55	73
Advanced 3	42	60
Total	116	159

As the table illustrates, the number of idioms in terms of types and tokens were different for the three books. To find out if these differences were significant, a chi-square was run (Table 2). As the table shows, significant differences were found ($\chi^2 = 230$, $df = 2$, $p = .000$).

Table 2. Chi-square for idiom types in the three textbooks

	Observed N	Expected N	Residual
19.00	19	38.7	-19.7
42.00	42	38.7	3.3
55.00	55	38.7	16.3
total	116		

Table 3 illustrates the results of the chi-square test intended to find out whether the differences between the textbooks in terms of the number of idiom tokens were significant. As the table shows, significant differences were found ($\chi^2 = 230$, df $= 2$, $p = .000$).

Table 3. Chi-square for idiom tokens in three textbooks

	Observed N	Expected N	Residual
26.00	26	53.0	-27.0
60.00	60	53.0	7.0
73.00	73	53.0	20.0
total	159		

To answer the second question, the next step was to find the frequency of idioms presented in the textbooks in the MICASE, BNC and Brown corpora. To achieve this, each idiom was searched online in the three corpora. Tables 4, 5, and 6 show the results of these analyses for each of the books, respectively.

Table 4. Advanced 1 idioms and their frequency bands across the three corpora

Corpus	Band 1 (above 50)	Band 2 (19-49)	Band 3 (2-18)	Band 4 (0-1)
MICASE	5	5	5	4
BNC	4	3	9	3
BROWN	4	2	8	5

Table 5. Advanced 2 idioms and their frequency bands across the three corpora

Corpus	Band 1 (above 50)	Band 2 (19-49)	Band 3 (2-18)	Band 4 (0-1)
MICASE	2	5	22	26
BNC	3	6	23	23
BROWN	3	7	25	20

Table 6. Advanced 3 idioms and their frequency bands across the three corpora

Corpus	Band 1 (above 50)	Band 2 (19-49)	Band 3 (2-18)	Band 4 (0-1)
MICASE	7	3	20	11
BNC	9	4	18	10
BROWN	6	3	20	12

Table 7. Distribution of idioms across four frequency bands

Textbook	Number of idioms	Band 1 (above 50)	Band 2 (19-50)	Band 3 (2-19)	Band 4 (0-1)
Advanced 1	19	8	6	3	2
Advanced 2	55	9	10	24	12
Advanced 3	42	13	10	16	3
Total	116	30	26	43	17

As could be understood from Table 7, from a total of 116 idioms, 30 fell in the first band, considered as common idioms in the corpora, and 26 idioms fell within the second band. Forty three idioms were in band 3 which are seen as the idioms with low frequencies. The last 17 fell within the zero-frequency band.

6. Discussion

According to Nation (2011), vocabulary should either be taught according to "frequency of occurrence, communicative need and complexity" (p. 386), or convenience. In case of the ILI textbooks, it seems that the latter approach, that of convenience, is adopted, as there does not seem to be any systematic selection based on frequency, or balance between the number of idioms selected for and included in each textbook. As our results showed, the idioms chosen for these advanced books were not very frequent in the corpora. According to Sinclair and Renouf (1988) and Leech (2001), among the most important principles for the inclusion of an item in a syllabus is the frequency of word forms, meanings and their inflections. With such low frequencies of occurrence as those found for the idioms in the ILI textbooks, it seems that none of these criteria were accounted for in the selection process. To overcome this disadvantage, the writers of these textbooks are advised to use word/idiom frequency lists to select appropriate and frequent idioms for inclusion in their materials. Another way to tackle this issue, emphasized by McCarthy (1998), is to raise students' awareness regarding idioms, and to create a context of use which is as interactive as possible, encouraging real life authentic usage rather than decontextualized memorization. As such, the problem of low frequency idioms can be compensated for by the rich context provided by the instructor or course book.

In the same line, Simpson and Mendis (2003) suggest holding workshops for students during which students are made aware of the nature of idioms and try to identify idioms from spoken contexts and examples from corpora with special emphasis on discourse markers, context clues and even glosses and paraphrases.

The significant differences found between the frequencies of occurrence of each idiom across the three corpora can be discussed in light of the fact that the three corpora used as references for the present study were different from several aspects. Such differences are inevitable, since no two corpora, even if created based on a similar linguistic body, will ever be identical. In addition, no matter how close the genres within the corpora are, there will always be formal and contextual differences caused by the speakers, the nature of the speech acts and the function of each selection.

Leech (2001), points out certain issues relevant to the present work that should be considered when using corpora. First, he mentions the difficulties of finding the right corpus in terms of relevance, size and students' needs. In other words, the corpus used for vocabulary teaching should be representative of the language to be included in books. Leech emphasizes the difficulty of providing a definition for the notion of representativeness; nevertheless, having a general understanding of the concept as "balanced samples of a wide range of texts and transcribed speech" (p. 6), in our case, it is possible to say that the scope of the corpora did not represent exactly the goals set by the course books, which might have resulted in low frequency counts of the target idioms.

Other issues pointed out by Leech (2001) are that even if a corpus is large enough, the variety of language it provides might not match with the students' needs (general English, ESP, EAP, etc.), or their levels, and some corpora, including the BNC, might only be useful for more advanced students. Accordingly, one must study each corpus carefully before analyzing them for the most frequent MWUs to include in any textbook.

Another relevant point, mentioned by Simpson and Mendis (2003), is that although representative of a large number (over 150) of speech events, with only 1.7 million words, MICASE is a relatively small corpus, and the frequency of any particular idiom in this corpus cannot expected to be too high. The same can also be said about the Brown Corpus with only 1 million words.

7. Conclusion and Implications

The results of the present study revealed that the ILI advanced books are significantly different in terms of both number of idiom types and tokens. Among the identified idioms in the three textbooks, only 25% had frequencies over 50 per million tokens. The textbooks did not show any specific pattern in terms of the number of idioms each contains, which is an indicator that the important issue of idiomaticity and frequency has been overlooked during the compilation process. Thirty two were found to have significantly different frequencies across the corpora which could be related to different features of each corpus such as date of compilation, size and English variety and/or genre.

The findings of the present study have two basic pedagogical implications. First, it seems that the idiomatic expressions in these textbooks need to be selected in a more systematic way and should be based on authentic language rather than the writers' intuition in order to increase their content representativeness. However, since revising the material or re-writing it from scratch is an extremely difficult task, as Simpson and Mendis (2003) suggested, the meanings of the already included idioms can be highlighted using multiple choice activities, examples from real contexts, or comparisons of idiomatic meanings of MWUs with literal presentations of the same meanings. They also suggest presenting different meanings of idioms which are determined by their specific context of use as well as the different parts of speech they might have in different sentences. Kennedy (2003) and Criado and Sanchez (2012) also suggests more frequent exposure to MWUs and collocations to further enhance their learning. Implicit internalization would also be maximized, he believes, if the word combinations are met frequently enough both in and out of the classroom.

Second, ESL teachers, especially those of low-level students, might want to refer to corpus-based lists of the most frequently used idioms when selecting idioms to teach in the classrooms, particularly when more objective data on frequency is easily available. Such consultation may help decrease the chance of having students work on idioms not useful to them at the time of instruction, and the students will no longer need to learn less frequent idioms which are hardly ever used in real life; instead, they will learn those which are most likely to be encountered in their future English usage. Another advantage, according to Leech (2001) is the convenience of using frequency data for usefulness measurments as opposed to other selection methods. To this goal, to create new materials, corpus-based lists of idioms and MWUs (for example Leech et al, 2001; Simpson & Mendis, 2003; Liu, 2003; Martinez & Schmitt, 2012) could be referred to before the material is written, so that only those idioms are included that have appeared in these high-frequency idioms lists.

A number of limitations need to be acknowledged and addressed regarding the present study. The first is the fact that the search for idioms in the corpora was based only on the headwords of each MWE, not all derivations of a word. A more detailed study will be needed to address each idiom in its several various and possible forms. Besides, in the current research, as Sinclair and Renouf (1998) have also emphasized, multiple meanings of idioms were not considered; headwords of idioms such as *take off* and *take out* and *call for* and *call upon* have different meanings which were not accounted for in this study.

The second limitation has to do with the extent to which the findings can be generalized beyond the sample textbooks studied. Before revisions are made, other studies will be required to find out whether the same pattern of idiom selection emerges for similar corpora and different ILI book levels. Applying similar research to other textbooks can also unveil possible patterns which can be useful for future textbook development.

References

Aijmer, K. (2009). *Corpora and Language Teaching*. Amsterdam: John Benjamins.

Barlow, M. (2000). *Corpus of spoken, professional American English* [CD-ROM]. Houston, TX: Athelstan.

Biber, D. (2006). *University Language: A Corpus-based Study of Spoken and Written Registers*. Amsterdam: John Benjamins Pages.

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

Campoy, M., Gea-valor, M., & Belles-Fortuno, B. (2010). *Corpus-based Approaches to English Language Teaching*. London: Continuum.

Celce-Murcia, M., & Larsen-Freeman, D. (1999). *The grammar book: An ESL/EFL teacher's course* (2nd ed.). Boston: Heinle & Heinle.

Criado, R., & Sánchez, A. (2012). Lexical frequency, textbooks and learning from a cognitive perspective. A corpus-based sample analysis of ELT materials. *Revista Española de Lingüística Aplicada (RESLA)*, *Special*

Issue, 77-94.

Ellis, R. (1985). *Understanding second language acquisition*. Oxford: Oxford University Press.

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

Francis, W. N., & Kucera, H. (1964). *A standard corpus of present-day edited American English, for use with digital computers*. Providence, RI: Brown University, Department of Linguistics.

Gardner, D., & Davis, M. (2007). Pointing out frequent phrasal verbs: A corpus-based analysis. *TESOL Quarterly*, *41*(2), 339-359.

Garside, R., Leech, G., & McEnery, A. (Eds.). (1997). *Corpus Annotation: Linguistic Information from Computer Text Corpora*. Longman, London.

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

Kennedy, G. (2003). Amplifier collocations in the British National Corpus: Implications for English language teaching. *TESOL Quarterly*, *37*(3), 467-487.

Kirk, J. M. (2000). C*orpora Galore: Analyses and Techniques in Describing English*. Amsterdam: Rodopi.

Leech, G. (2001). *The Role of Frequency in ELT: New Corpus Evidence Brings a Re-appraisal*. Retrieved September 10, 2014, from http://www.lancaster.ac.uk/fass/doc_library/linguistics/leechg/leech_2001.pdf

Leech, G., Rayson, P., & Wilson, A. (2001). *Word Frequencies in Written and Spoken English*. London: Longman.

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., & Schmitt, N. (2012). A Phrasal Expressions List. *Applied Linguistics*, *33*(3), 299-320.

McCarthy, M. (2004). *Touchstone-from corpus to course book*. Edinburgh: Cambridge university press.

McCarthy, M. J. (1998). *Spoken language and applied linguistics*. Cambridge: Cambridge University Press.

McEnery, T., & Xiao, R. (2010). What corpora can offer in language teaching and learning. In E. Hinkel (Ed.), Handbook of Research in Second Language Teaching and Learning (Vol. 2, pp. 364-380). London & New York: Routledge.

Mindt, D. (1996). English Corpus Linguistics and the Foreign Language Teaching Syllabus. In J. Thomas, & M. Short (Eds.), *Using Corpora for Language Research*. London: Longman.

Minugh, D. (2002). *The college idiom: Idioms in the COLL corpus*. Stockholm: University of Stockholm.

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

Nation, I. S. P. (2011). Research into practice: Vocabulary. *Language Teaching*, *44*(4), 409-411.

O'Keeffe, A., McCarthy, M. J., & Carter, R. A. (2007). *From Corpus to Classroom: Language use and language teaching*. Cambridge: Cambridge University Press.

Shin, D., & Nation, I. S. P. (2008). *Beyond single words: The most frequent collocations in spoken English*. Oxford: Oxford University Press.

Simpson, R. C., Briggs, S. L., Ovens, J., & Swales, J. M. (2002). *The Michigan corpus of academic spoken English*. Retrieved July 14, 2011, from http://www.hti.umich.edu/m/micase

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

Sinclair, J., & Renouf, A. (1988). A lexical syllabus for language learning. In R. Carter, & M. McCarthy (Eds.), *Vocabulary and Language Teaching* (pp. 140-160). Longman.

Sporleder, C., Linlin, L., Gorinski, P., & Koch, X. (2010). *Idioms in context: The IDIX corpus*. Paper presented at he seventh international conference on Language Resources and Evaluation (LREC), Valletta, Malta.

Trebits, A. (2009). *The most frequent phrasal verbs in English language EU documents – A corpus-based analysis and its implications* (Unpublished manuscript). Eotvos Lora´nd University, Budapest, Hungary

Walker, C. (2011). How a corpus-based study of the factors which influence collocation can help in the teaching of business English. *English for Specific Purposes*, *30,* 101-112.

Appendix

Idioms included in each book

Advanced 1 Idioms	put out, stressed out, used to, point out, go on, try out, break down into, take place, make up, cut off, find out, turn out, bark up, bound up, look up, as well as, take out, have to, fall into
Advanced 2 Idioms	kick off, break for, run together, give up, in short, look for, get down to, come down to, in terms of, stack up, own up, down payment, carry on, break up, run into, bring up, cope with, fall out, burst into, go by, set forth, hold back, tie up, get back, face up, make up, carry out, die out, tip off, sniff out, order out, set off, watch out, find out, turn over, hook up, act up, pull over, ought to, come over, take off, book club, help out, turn into, on the one hand, spell the end, get on to, call for, seek out, lip service, call out, keep up with, flight out
Advanced 3 Idioms	go on, set out, cover up, have to, walk by, pluck up, get through, run through, in order to, keep up, such as, drop out, due to, lay out, cry out, run away, try out, find out, live up to, carry out, come by, light up, shut down, turn out, pave the way, brush off, block out, come off, account for, point out, set up, take over, take off, break up, get together, call on, turn down, get along with, look forward to, out-of-date, bona fide

Copyrights