The Most Frequent Idioms Used in Contemporary American English:
A Corpus-based Study

Elaheh Rafatbakhsh [1], Alireza Ahmadi [2*]

[1] *PhD Candidate of TEFL, Department of Foreign Languages and Linguistics, Shiraz University, Shiraz, Iran*
[2] *Associate Professor of TEFL, Department of Foreign Languages and Linguistics, Shiraz University, Shiraz, Iran*

Abstract: As a fascinating and colorful part of English language, idioms highly affect fluency, but they are quite difficult to teach and learn, and they have often been neglected particularly in ESL/EFL settings. Considering the large number of English idioms, corpus linguistics can be of great benefit in prioritizing materials in language classrooms based on the frequency information. Accordingly, the present corpus-based study aimed at identifying the most frequent idioms in English language by analyzing the data coming from the Corpus of Contemporary American English (COCA), composed of more than 520 million words. The study involved writing a special script using Hypertext Preprocessor (PHP) language which resulted in the development of five idiom lists, each containing 50 most frequently used idioms in each one of the five genres in COCA including academic, fiction, spoken, newspaper, and magazine along with their frequency of occurrences. Comparison was then made across the mentioned five genres. It was found that the spoken genre included more idioms whereas the academic genre was the least idiomatic. Furthermore, various levels of overlap was found among different genres. The least and the highest levels of overlap was found between the academic and fiction genres and between the magazine and newspaper genres respectively. The academic genre had more overlap with the newspaper and magazine genres. The findings can benefit EFL materials developers, teachers, and learners in recognizing and including frequently-used authentic idioms in language classrooms and textbooks.

Keywords: Corpus of Contemporary American English (COCA), EFL, Frequency List, Idioms, Materials Development.

---

[*] **Corresponding Author.**

Authors' Email Address:
[1] **Elaheh Rafatbakhsh** (e.rafatbakhsh@shirazu.ac.ir), [2] **Alireza Ahmadi** (arahmadi@shirazu.ac.ir),

Introduction

Language is not just vocabulary and grammar as it was formerly believed; in fact, it consists of multi-word prefabricated chunks or in other words "formulaic language" (Lewis, 1993). Formulaic language, including expressions such as phrasal verbs, compounds, idioms, and collocations, has an important role in fluency and also motivating the learners (Schmitt, 2000). According to Fraser (1970), an idiom is a "constituent or series of constituents for which the semantic interpretation is not a compositional function of the formatives of which it is composed" (p. 22). In line with that definition, Gramley and Patzold (2003) defined idiom as a "complex lexical item which is longer than a word form but shorter than a sentence and which has a meaning that cannot be derived from the knowledge of its component parts" (p. 55).

In general, idioms are basically fixed expressions, understood by native speakers of a language, whose characteristics cannot normally change and the meaning cannot always be guessed from the meaning of the idioms' component words, hence they can get tricky for the learners at times (Berman, 2000). Idiomatic expressions often indicate and reflect social norms, beliefs, attitudes, and emotions. Learning idioms and other fixed expressions represents learning a culture (Crystal, 1997, Glucksberg & McGlone, 2001, Ovando & Collier, 1985). Idioms are what give English language variety and imagination and without them it would be bookish and stilted (Cooper, 1999).

Three main reasons for learning idioms by L2 learners are proposed by Boers, Eyckmans, Kappel, Stengers, and Demecheleer (2006). On top, native-like proficiency is achieved in time through learning idioms; secondly, learning idioms in chunks can help learners to retrieve them from their memory with fewer hesitations, and finally, it can give the learners fluency especially in real-time situations and authentic communications. Idioms and metaphorical expressions are highly believed to be crucial in building fluency (Cain, Towse, & Knight, 2009; Lim, Ang, Lee, & Leong, 2009; Teodorescu, 2015), therefore language learners need to be aware of these chunks and their functions. Idioms are not quite easy for L2 teachers to teach; they are pretty difficult for L2 learners to learn, and helping learners in idiom acquisition has always been a challenge (Liu, 2003). As a solution to the problem of teaching and learning idioms, applied linguistics, and corpus linguistics in particular, can be of great help in identifying frequency and patterns of idiom use, and therefore giving the idioms priorities in teaching and learning in L2 contexts (Liu, 2003).

Not all textbooks and materials in L2 education have adopted corpus-based approaches when it comes to vocabulary and idioms. Considering the large bulk of material, both linguistic and non-linguistic, and the short time language learners often have for learning a language, it is clearly more crucial to first focus on the most frequent and real-life parts of a language rather than less practical ones. Corpus linguistic have made it possible to decide what is more or less important to use in language classrooms based on their frequency information (Biber & Reppen, 2002). This research reports a corpus-based study of dictionary idioms with the aim of identifying the most frequently used idioms in the spoken, academic, fiction, newspaper, and magazine genres.

Background

The significance of learning idioms has been emphasized by many scholars and researchers as it provides acquaintance with the target language culture and helps develop communicative competence, proficiency, and fluency (Bardovi-Harlig, 2002; Fernando, 1996; Liu, 2008; Moon, 1998; Schmitt, 2004; Thyab, 2016; Wood, 2002; Wray, 2000). Boers (2013) believes that teaching and learning idioms are important and significant in language learning, since, firstly, lack of knowledge of idioms can cause serious comprehension problems and misunderstandings in many contexts even if they are rich in clues; and secondly, the use of idioms and figurative idioms in particular, is not as infrequent as it has been assumed. In the same line, Maisa and Karunakaran (2013) conducted an exploratory research on the importance of teaching idioms to ESL students from teachers' perspective. The results showed that teachers believed teaching idioms to undergraduate students as an integral part of vocabulary teaching, leads to more fluent speaking and writing. Moreover, it is quite beneficial to include idioms in dialogues, readings, and stories in the curriculum.

Other researchers have tried to identify the most frequent English idioms in different corpora in order to give priorities to teachers and learners. However, searching for idioms in corpora is a difficult and complex process as idioms consist of various parts and they might even spread over the whole sentence (Busta, 2008). For instance, in order to be able to search for idioms, Baddorf and Evens (1998) used a small program suite to search corpus files for the list of phrases. They searched a list of 30 phrases and idioms and their syntactic variants from Collins English Dictionary in three corpora, Wall Street Journal (WSJ) corpus (47,456,421 words), Dictionary of Old English (DOE) corpus (27,944,329 words), and the

corpus of Gutenberg (41,588,806 words). Before searching, they transformed the phrases in order to be able to find all different variants of each phrase.

Another main corpus study was carried out by Moon (1998) in which 6776 commonest British and American English Fixed Expressions including Idioms (FEI) in a premade database were searched for in the Oxford Hector Pilot Corpus (OHPC). The findings yielded information on overall frequencies and distributions and explanations were provided on lexical and grammatical form, variation, ambiguity, polysemy, metaphor, discoursal functions, evaluation and interactional perspectives, and cohesion in FEIs. Conclusions drawn from this study showed that further studies are required to create a more accurate image of the expressions. Furthermore, the study suggested that more revisions of existing models and descriptions should be made and the importance of the roles FEIs play in discourse should not be underrated.

In another study carried out by Liu (2003), the problems of intuition-based teaching, choosing materials on English idioms, and also incorrect descriptions of the meaning and use of some of these idioms were addressed. In order to do so, Liu conducted a corpus-based study and analyzed the frequently used idioms in three contemporary spoken American English corpora: Corpus of Spoken, Professional American English (Barlow, 2000); Michigan Corpus of Academic Spoken English (Simpson, Briggs, Ovens, & Swales, 2002), and Spoken American Media English (compiled by Liu). The idioms were identified using Fernando's (1996) three categories (pure, semiliteral, and literal). Phrasal verbs were also included in this study as a lot of them are fixed and they have nonliteral or semiliteral meanings. Four major contemporary English idiom dictionaries and three English phrasal verb dictionaries including Cambridge International Dictionary of Idioms (1998) and Cambridge International Dictionary of Phrasal Verbs (1997), Longman American Idioms Dictionary (1999), NTC's American Idioms Dictionary (Spears, 2000) and NTC's Dictionary of Phrasal Verbs and Other Idiomatic Verbal Phrases (Spears, 1993), and Oxford Idioms Dictionary for Learners of English (2001) and Oxford Phrasal Verbs Dictionary for Learners of English (2001) were consulted for idiom identification. Difficulty of an idiom is directly connected to how literal that idiom is. Therefore, to decide on the difficulty of an idiom, a fairly literal expression was considered an idiom if it was listed in two idiom dictionaries or two phrasal verb dictionaries. The results of the study provided four lists of the most frequently used idioms and their use patterns. The research also manifested that the use of idioms in teaching and reference materials is inadequate in terms of item selection, meaning,

explanation, and the examples provided and some suggestions were presented for improvements.

Simpson and Mendis (2003) also carried out a corpus-based study of idioms in academic speech. They investigated a specialized corpus of 1.7 million words of academic discourse, the Michigan Corpus of Academic Spoken English. In order to identify an idiom in this research, three criteria of compositeness or fixedness (having fixed lexical units that cannot be simply substituted for and replaced), institutionalization (conventionalization of a novel expression), and semantic opacity (not having transparent meaning) were used. This study revealed a list of frequent idioms occurred in academic speech in addition to their functions. They concluded that the use of idioms in academic speech is not a rare phenomenon. In addition, the use of idioms is not related to the socio-interactional roles of the speakers, rather it is a feature of the individuals' idiolects. This research also suggests the use of corpus in teaching and learning idioms as it can provide both teachers and learners with authentic contexts and examples of idiom use as well as considering sociopragmatic and interactional features of idioms rather than their immediate context. Two idiom lists were presented in this study, first the idioms which are particularly useful for English for academic purposes curricula, and idioms that occurred four or more times in the Michigan Corpus of Academic Spoken English (MICASE).

Furthermore, Grant (2007) assessed frequent spoken figurative idioms in order to help ESL/EFL teachers decide on which idioms are more useful to teach to students first. In this study, the frequent figurative idioms identified in two sources of spoken American English (academic and contemporary) and in spoken British English were compared by searching the spoken part of the British National Corpus (BNC). The comparison also included figuratives identified as frequent in two British idiom dictionaries. The idiom lists from Liu's (2003) and Simpson and Mendis's (2003) researches were employed in this study. The results were presented as tables manifesting the comparison of frequencies of figuratives in MICASE and the spoken BNC as well as in spoken American and British English. According to the findings, idioms do occur in academic genre, however, their distribution does not appear to be predictable and using them is connected to individual speakers' idiolects rather than other factors.

The corpus search of idioms so far has been limited to small number of idioms including only some specific types of idioms such as core idioms or figuratives. Additionally, these studies have chosen only particular domains such as the academic or spoken section of

a corpus. Not much research has been conducted to investigate a wider range of idioms in a broader and larger corpus such as the Corpus of Contemporary American English with different genres and domains. Searching a large number of idioms along with their variations in large corpora can be quite difficult and time consuming. Thus, the purpose of the present corpus-based study was to fill the research gap in the literature by searching all the idioms of the Oxford Dictionary of Idioms in the large Corpus of Contemporary American English (COCA) using a script especially written for this purpose by a professional programmer since the existing concordancers had difficulty searching this great range of idioms and their variations in a corpus as large as COCA. Overall, the study aimed at searching for and comparing the idioms in five different genres of spoken, academic, newspaper, fiction, and magazine in COCA which represents the biggest corpus available so far for the English language.

Methodology

Since this was a descriptive study and the purpose was to identify the most frequent idioms of English in COCA, the quantitative design was employed for extracting each idioms frequency.

Corpus

The research was based on the Corpus of Contemporary American English (COCA) which is composed of more than 520 million words in 220,225 texts, including 20 million words each year from 1990 to 2015. COCA was created by Mark Davies (2008), Corpus Linguistics professor at Brigham Young University. Currently, COCA is the most recent, comprehensive and balanced corpus of English language that exists. This corpus is divided evenly in five genres of spoken, fiction, popular magazines, newspapers, and academic journals for each year and also overall. Each genre comes from various authentic sources. The genre of spoken consists of 109 million words (109,391,643) which are transcripts of unscripted conversations from more than 150 different television and radio programs. The genre of fiction which has 105 million words (104,900,827) is from short stories and plays from literary magazines, children's magazines, popular magazines, movie scripts, etc. Popular magazine genre includes 110 million words (110,110,637) from about 100 magazines with a balanced mix of specific domains such as news, health, home and gardening, women, financial, religion, sports, etc. For newspapers, 106 million words (105,963,844), a good mix of various sections

such as local news, opinion, sports, financial, etc. Finally, the genre of academic journals with 103 million words (103,421,981), is from about 100 peer-reviewed journals covering the entire range of the Library of Congress classification system. It should be noted than the purchased corpus of COCA has 95% of the whole data and 5% is removed by the owner due to reasons of copyright which may have slightly affected the search results. Table 1 summarizes COCA, its genres, and the number of words in each from 1990 to 2015.

Table 1. *Different Genres and their Size in COCA (1990-2015)*

Genre	Spoken	Fiction	Magazine	Newspaper	Academic
number of words	107,973,088	103,418,530	109,014,187	104,618,087	103,295,116

Idiom Dictionary

As the main source of the idioms, the Oxford Dictionary of Idioms, 2nd ed., was chosen because it was the latest dictionary of idioms available in digital format for this research. In addition, this dictionary with about 5000 American and British English idioms gives a combination of definitions, explanations and illustrative quotations where needed and provides a thorough picture of how an idiom is used. All the idioms of this dictionary were included as the research was not limited to any specific types of idioms.

Data Collection Procedure

Idioms Transformation and Coding Procedure: To start with, all idiom entries were extracted from the digital version of the Oxford Dictionary of Idioms manually and a list of 4986 idioms was created. Normally searching corpora for frequency data is carried out by concordancers. Concordancers such as WordSmith Tools (Scott, 2012) and AntConc (Anthony, 2009) are computer programs for text analysis usually used in corpus linguistics to retrieve alphabetically or other sorted lists of linguistic data from a corpus. However, they face some limitations when dealing with a large corpus and a big list of data. As a solution, Anthony (2009), Biber, Conrad, and Reppen, (1998), Gries (2009), and Weisser (2009) proposed that for the best outcome, it is better and more convenient for the corpus linguists to develop their own tools for text analysis based on their specific needs and purposes. Therefore, due to the large number of idioms and the big size of COCA, a script was written in PHP by a professional computer programmer in order to find out the frequency of each idiom in the corpus. The script broke the big COCA into smaller pieces and then used regular expressions to search the number of occurrences of each idiom in COCA. All the idioms were

prepared to be used for programming manually one by one to search for various forms of each idiom. The formulas were written as follows:

1. All the main verbs of the idioms that could change forms depending on the context were capitalized (e.g. the idiom "go ape" was rewritten as "GO ape").

2. The words in some idioms which were not fixed and could be replaced by other words, were changed into an asterisk symbol. These words include: Possessive adjectives, subject and object pronouns, something, someone, somebody, one, etc. (e.g. the idiom "tied to someone's apron strings" was transformed to "tied to * apron strings").

3. Additionally, after the observation of different idioms in real contexts, asterisks were coded in a way to include from no words up to three words. For instance, the idiom "tied to someone's apron strings" could be "tied to HER apron strings", "tied to MOTHER 'S apron strings" or "tied to HIS MOTHER 'S apron strings" or in some cases it gives more space for possible adjectives and adverbs in the idioms. For instance, the idiom "be one's own man (or woman or person)" might be "be HIS OWN man" or "be VERY MUCH HIS own man".

4. In the idioms that had two or more alternative words, the symbol "|" was used to separate the nodes (e.g. the idiom "as clear (or sound) as a bell" was rewritten as "as clear|sound as a bell.").

5. Contractions such as "'s, n't, 're, etc." in some idioms were separated by a space as written in COCA (e.g. The idiom "big girl's blouse" was changed to "big girl 's blouse").

6. Some idioms were shortened if possible, to ease the search and also to cover more related idioms (e.g. the idiom "a bolt from the blue" was rewritten as "bolt from the blue").

7. Some idioms with two different possible forms that could not be merged in one had to be written as two entries to be added up later (e.g. the idiom "arrow of time (or time's arrow)"). Therefore, the total number of idioms increased to 5083.

8. Finally, the words such as color, favorite, etc. with two possible spellings were identified and both forms were added to the search (e.g. the idiom "with flying colours" were rewritten as "with flying colours|colors").

All English verbs along with their forms (past tense, past participle, third person, and gerund) were added to the coding process. Moreover, for the verbs such as have, to be, and modal verbs, their negative forms as well as their contractions were also included in the verb list. For instance, for the verb have, the following were added and searched in the corpus: have, has, having, had, 've, 's, 'd, haven't, hasn't, hadn't, have got, has got, haven't got, hasn't got, hadn't got, have not got, has not got, and had not got.

Searching the Corpus and Making the Lists: After the creation of the computer program and before the main search, to test the accuracy of the developed system, 20 idioms were randomly chosen and searched via the system and also COCA's online concordancer and the frequency results where then compared. The frequencies received by the system and the ones obtained from the website's online concordancer were the same. Besides improving the accuracy of the system, the searching program was also optimized for more efficient and rapid results. Consequently, the system searched all the idioms, their tokens and variations in the entire corpus in short amount of time.

The system provided a detailed spreadsheet of statistics with 5083 rows and 127 columns. Each cell represented frequency information about each genre (academic, fiction, spoken, news, and magazine) and each year (1990-2015) for each idiom, for instance, academic-1990, academic-1991, academic-1992, etc. The cells were summed up for each genre separately and the results were added to the table. After the search, to make the frequency list of each genre, first the results were sorted and then the most frequents of each list were manually confined by reevaluating some of the idioms using COCA's website to have more precise data. Some idioms were searched again using part of speech (POS) tagging feature of the website. For instance, the idiom close to (or close on) was searched again as "close to (or close on) + number" to be in line with its idiomatic meaning which is related to an amount. Other idioms such as 'on it', 'for all', 'out for', and 'out with' were removed as it was quite impossible to limit the search results using part of speech tagging to eliminate non-idiomatic use of phrases. The frequencies of the top idioms that included asterisks were also checked again in order to omit irrelevant words that had substituted an asterisk. In addition, some entries of the dictionary were similar in wording and/or meaning, therefore, they were either merged or separated. For instance, the frequency of idiom "the rest is history" was subtracted from the frequency of the idiom "be history" as they were similar in words but different in meaning.

Finally, 50 most frequent idioms of each genre were sorted and their frequencies were calculated per million. Idioms repeated in all five genres were identified and reported. Furthermore, descriptive statistics were computed and analyzed using SPSS (Statistical Package for the Social Sciences) version 24. And in order to compare the genres, the percentage of common idioms across genres were calculated.

Findings

Effort was made to cover all the possible forms of each idiom, however, because of the nature of idioms, the full cover cannot be claimed. To find the most frequent idioms of each genre, after the computer search, some of the most repeated idioms were manually checked and searched again in the website of COCA and the results were modified if necessary. Therefore, some idioms were eliminated or replaced from the top idioms list of each genre. Only the first 50 most frequent idioms in each genre are reported in this paper.

The frequency results from the spoken section were sorted based on their occurrences per million and the results are demonstrated in Appendix A. For instance, the idiom "every last (or single)" was the most frequent idiom in the spoken genre with a frequency of 33.61 per million words. Both "every last" and "every single" were searched in the COCA by the system and summed up automatically. "Behind closed doors" is another frequent idiom repeated ٤.٥٥ times per million words in this genre. One example sentence was randomly chosen for each idiom from COCA to illustrate the use of these idiom in the context. The sources and the dates of the example sentences are also mentioned:

(1) (TED Radio Hour, 2015): Maybe these conversations will remind us what's really important, and maybe it will help us recognize that simple truth that every single life matters equally and infinitely.

(2) (Fox, 2008): Well, it's hard for me to know exactly what's going on behind closed doors.

The search also resulted in the second list, which is the 50 most frequently used idioms in the fiction genre of COCA along with their approximate frequency per million (see Appendix B). The most frequent idiom of this list is "big deal" with a frequency of about 13.18 per million. Another frequent idiom in this list is "do someone a favour" with a frequency of nearly 7.12 including all the variations and forms of this idiom. Hence, this frequency includes all the forms of the verb "do", along with the substitutes of the word "someone" which starts from no word up to three words, and finally the word "favour" was searched with another spelling, "favor", as well. One random example for each idiom together with their sources and years in the fiction genre are brought here.

(1) (Postmortem, 1990): It's a fact we've been together, but big deal?

(2) (the movie "In the Valley of Elah", 2007): I'm just doing a favor for a neighbor; her boy got in a little trouble.

Appendix C depicts the next list which is the 50 most frequent idioms in the magazine

genre sorted by their frequency of occurrence per million. The most repeated dictionary idiom in this genre is "the bottom line" that was employed about 11.05 per million throughout the mentioned context (including all forms of the verb "to be" together with their negative forms). Another example of a frequent idiom in this list is "fall short (of)" which has been repeated 4.94 times per million words. This idiom was searched in the corpus with all forms of the verb "fall" and without the preposition "of".

(1) (Popular Science, 2014): The bottom line is storage isn't the problem: It' s our ability to record and retrieve data that is.

(2) (America, 2011): M. Cathleen Kaveny, in "Defining Feminism", presents both sides of the issue but falls short in pointing out the uniqueness of each person regardless of gender.

Moreover, the 50 most popular idioms in newspaper were sorted based on their approximate frequency per million (see Appendix D). The idiom "the bottom line" is also the most repeated in this genre, repeated 12.03 times per million. "Come of age" is another frequent idiom in this list, 3.43 times per million, which includes all the verb forms of "come". Random examples for these two idioms are also provided below.

(1) (St Louis Post-Dispatch, 2014): In this business, the bottom line is production.

(2) (Associated Press, 2003): They came of age on city streets from Los Angeles to New York.

Finally, the 50 most frequently used idioms in academic genre were brought together and sorted based on their approximate frequency per million (see Appendix E). Based on the results, "in the long run (or term)" were repeated the most among the idioms in this genre, 9.56 times per million words. Another example idiom of this category is "open the door to", repeated 2.46 per million including all verb forms of "open".

(1) (Journal of International Affairs, 1992): Though the discipline of SAPs may lead to stronger economies in the long run, in the short-term their primary impact is to reduce governments' ability to meet their citizens' needs.

(2) (Agricultural Research, 2000): That lowers the cost of producing seed and opens the door to wider use of Indian rice grass in the West.

Of all the 50 most frequently used idioms, only 6 idioms were in common among all the five lists (see Table 2). From among the 4986 idioms, about 550 (11.03%) had zero occurrences in the whole corpus of COCA. Idioms such as "go like a bomb", "set by the ears", "an itching palm", "play a blinder", and "meet trouble halfway" were among them.

Table 2. *The Most frequently Used Idioms Common in all the Five Genres*

close to (or close on)	come (or spring) to mind
if anything	every last (or single)
for free	on someone's mind

Table 3 summarizes the descriptive statistics about the five genres. As indicated, the highest mean is related to the spoken genre with a mean of 8.14 and the lowest mean is related to the genre of academic with a mean of 2.93 The difference in the means of the frequencies is quite noticeable. This indicates that the genre of spoken English included more idioms; in other words, it was more idiomatic. In the same way, the academic language which tends to be more formal was less idiomatic and used more direct and scientific rather than figurative language. In addition, fiction, with a mean of 6.50 appeared to be more idiomatic than newspaper and magazine genres since the language of fiction is more likely to be less formal than the language used in newspapers and magazines.

Table 3. *Descriptive Statistics about the 50 Most Frequently-used Idioms in the Five Genres*

	Minimum	Maximum	Sum	Mean	Std. Error	Std. Deviation
spoken	3.34	33.61	407.17	8.14	1.00	7.10
fiction	3.67	13.18	324.79	6.50	.39	2.74
newspaper	2.61	12.30	248.24	4.96	.33	2.30
magazine	3.00	11.05	243.56	4.87	.27	1.92
academic	1.37	9.56	146.26	2.93	.23	1.61

The most frequent idioms were further compared to see whether the same idioms may be of frequent use across genres. Table 4 depicts the findings in this regard. As indicated, the academic and fiction genres had the least number of common idioms (20%) while the magazine and newspaper genres had the highest number of common idioms (74%). The academic genre had more overlap with the newspaper and magazine genres. This might be due to the fact that all of these three genres were written language and similarities among them can be expected. On the other hand, the level of commonality was lower between the academic genre and fiction and spoken. This is well acceptable as these two genres are not formal ones. Spoken in many cases is informal and fiction does not lend itself into the conventions of academic writing. In addition, the spoken genre had more common idioms with newspaper and magazine genres. This can be explained since the spoken genre also includes transcripts of conversation from different TV and radio programs inclusive of a lot of news programs besides its informal sections. For the same reason, the newspaper genre had the highest level of similarity with magazine and spoken genres; and also magazine genre

had the most common idioms with newspaper and spoken. These three genres seemed to be alike in terms of the idioms used as they all comprise and present some sort of news. Finally, the fiction genre had more similarity with the magazine and spoken genres comparing to the newspaper and academic genres. This seemed to be owing to the fact that both genres of spoken and magazine involve both formal and informal languages depending on the TV, radio program, or the magazine transcriptions. However, the language of newspapers and academic are mainly formal, direct, and less idiomatic unlike fiction.

Table 4. *Percentage of Common Idioms across Genres*

Genre	spoken	fiction	newspaper	magazine	academic
spoken		40%	64%	56%	34%
fiction	40%		34%	42%	20%
newspaper	64%	34%		74%	56%
magazine	56%	42%	74%		52%
academic	34%	20%	56%	52%	

In line with the results of the corpus search carried out by Grant (2007), Liu (2003), and Simpson and Mendis (2003), corpus-based studies of idioms are quite beneficial and crucial as idioms are not as rare as imagined before when considered as a whole. A more detailed comparison of the findings of this study and the previous similar ones is not quite possible since the choice of idioms searched in each study as well as the approach of searching for the idioms were different.

Conclusion and Implications

The corpus-based search of all the idioms of the Oxford Dictionary of Idioms in the largest freely-available corpus of English, COCA, has resulted in the development of five lists of the most frequent idioms in five different genres including spoken, fiction, magazine, newspaper, and academic. The 50 most frequently used idioms of each genre were reported in this paper along with their frequencies of occurrence per million words (see Appendixes 1-5). The 6 common idioms in all the five genres were also identified and presented. Additionally, based on the results of the descriptive statistics of the idioms' frequencies, the genre of spoken had the highest number of idioms among all, then fiction, newspaper, magazine, and academic respectively. Finally, a comparison was made across genres to demonstrate the percentages of common idioms in the five categories. The results revealed that the highest number of common idioms was between the genres of magazine and newspaper while the fiction and

academic genres had the least number of common idioms comparing to the other genres. Taken as a whole, the magazine, newspaper, and spoken genres tended to have more similar idioms.

Such corpus-based studies can be valuable as it provides information on what is actually used in the target language by the native speakers and most importantly how different elements of language are embedded in various contexts.

The findings of the present study, idiom lists and the frequencies, are not much comparable to the previous similar studies (Grant, 2007; Liu, 2003; Moon, 1998; Simpson & Mendis, 2003), since all types of idioms recorded in the thematic list of the Oxford Dictionary of Idiom were searched in all five genres of COCA, whereas former studies have focused on certain types of idioms such as core or figurative idioms in only one genre such as academic spoken English or spoken English.

These types of studies are based on frequency and range of occurrence in authentic language and should now replace authors' intuitions in developing materials used for teaching and practicing idioms. Therefore, the result of the present study can be beneficial to English teachers, learners, and materials designers and developers. The teaching and reference materials on idioms should be more rigorous and based on authentic language. To select materials, frequency and authentic use should be considered among other factors which is now possible to obtain using corpus-based researches. Giving priorities to the more frequent idioms for instance, is of great importance for teachers and learners. The presented lists can help materials developers to choose idioms based on their frequencies and also considering the genres they are involved in. Furthermore, examples provided for the idioms can switch to real examples from the corpus instead of made-up contrived sentences created only for learners.

Boers (2011) proposed Cognitive Semantic (CS) approaches as a great way of teaching figurative phrases such as idioms after testing Lakoff and Johnson's (1980) Conceptual Metaphor Theory (CMT). In this respect, classification of idioms based on metaphors, their source domains or origins has proved to be quite effective for learning and retention of the idioms. Making use of emotions is a popular example of conceptual metaphors with which idioms can be categorized. For instance, idioms that represent happiness such as "on top of the world", "over the moon", and "on cloud nine" should be presented together.

An alternative way to present the idioms is to employ pictures besides verbal education to assist learners' comprehension. According to Boers, Piquer Píriz, Stengers, and Eyckman's

(2009) investigations, pictorial education of idioms can foster the retention of idioms and make them more plausible for the learners. The appropriate use of photographs or drawings, if applicable, contributes to idiom learning and recollection. However, drawing the students' attention to the images should be in addition to verbal explanations of the certain formal features to enhance retention of both semantic and syntactic aspects. For instance, the idiom "off the hook" as one of the most frequently used one in the spoken genre of COCA, can first be taught verbally focusing on the formal aspects as well as the semantics, and then images can be added to the instructions to help the formation of mental pictures for the learners.

Another technique of teaching idioms in EFL/ESL contexts is to connect them to their similar idioms or concepts in the learners' native language if present. Similar idioms can be found in different languages and cultures and this can be a great topic for teachers to maneuver on. For instance, the idiom "cat got someone's tongue" is almost the same in Persian language, therefore, highlighting this similarity can motivate learners and aid them to store and then use it more easily.

Additionally, Simpson and Mendis (2003) suggested two types of exercises to use in the classroom which help expanding the knowledge of idioms. Using excerpts of authentic language from corpora and asking the learners to guess the meaning of the idioms is an example of such exercises. However, the extracts should be selected with care, i.e. enough contextual clues must be present to avoid comprehension difficulties. Below is an excerpt from the magazine genre of COCA which can be a good practice example of this type,

(MAG: Ms., 2006): Dr. Caruso said. It may become difficult to concentrate on

two things at once. The ability to multi-task or think on your feet may diminish.

They also believed multiple-choice exercises on the idioms' meanings can be quite helpful. As the question sample presented by Simpson and Mendis (2003) shows, the learners have to choose the best definition that matches the idiom in the stem.

"keep tabs on

a) agree with something

b) continue at the same pace

c) observe or record carefully

d) keep a secret (p. 440)"

Overall, the number of idioms used in authentic language has proved to be more than it was once imagined based on the advances in corpus linguistics and also in technology. Furthermore, lack of knowledge of idioms may result in comprehension difficulties for

ESL/EFL learners. Therefore, they should be included in the curricula and the classrooms using the mentioned approaches and techniques. The frequency results of such studies can be of help in deciding on the materials more systematically.

References

Anthony, L. (2009). Issues in the design and development of software tools for corpus studies: The case for collaboration. In P. Baker (ed.), *Contemporary corpus linguistics,* (pp. 87-104). London, UK: Continuum Press.

Baddorf, D. S., & Evens, M. W. (1998). Finding phrases rather than discovering collocations: Searching corpora for dictionary phrases. *Proc. of the 9th Midwest Artificial Intelligence and Cognitive Science Conference (MAICS-98)*, 110-116.

Bardovi-Harlig, K. (2002). A new starting point? Investigating formulaic use and input in future expression. *Studies in Second Language Acquisition*, *24*(2), 189-198.

Barlow, M. (2000). Corpus of spoken, professional American English [CD-ROM].

Berman, A. (2000). Translation and the Trials of the Foreign. In L. Venuti (Ed.), *The Translation Studies Reader* (pp. 284–297). London: Routledge.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics*. Cambridge: Cambridge University Press.

Biber, D., & Reppen, R. (2002). What does frequency have to do with grammar teaching? *Studies in Second Language Acquisition*, *24*(2), 199-208.

Boers, F. (2011). Cognitive semantic ways of teaching figurative phrases: An assessment. *Review of Cognitive Linguistics*, *9*(1), 227-261.

Boers, F. (2013). Cognitive linguistic approaches to teaching vocabulary: Assessment and integration. *Language Teaching*, *46*(02), 208-224. https://doi.org/10.1017/S0261444811000450

Boers, F., Eyckmans, J., Kappel, J., Stengers, H., & Demecheleer, M. (2006). Formulaic sequences and perceived oral proficiency: Putting a lexical approach to the test. *Language teaching research*, *10*(3), 245-261.

Boers, F., Piquer Píriz, A. M., Stengers, H., & Eyckmans, J. (2009). Does pictorial elucidation foster recollection of idioms? *Language teaching research*, *13*(4), 367-382.

Busta, J. (2008). Computing idioms frequency in text corpora. *Proceedings of Recent Advances in Slavonic Natural Language Processing*, Brno, Czech Republic. Masaryk University, 71-74.

Cain, K., Towse, A. S., & Knight, R. S. (2009). The development of idiom comprehension: An investigation of semantic and contextual processing skills. *Journal of Experimental Child Psychology*, *102*(3), 280-298.

Cooper, T. C. (1999). Processing of idioms by L2 learners of English. *TESOL Quarterly*, *33*(2), 233-262.

Crystal, D. (1997). *English as a global language*. Cambridge: Cambridge University Press.

Davies, M. (2008). Corpus of Contemporary American English (http://corpus.byu.edu/coca). *Brigham young university*.

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

Fraser, B. (1970). Idioms within a transformational grammar. *Foundations of language*, 22-42.

Glucksberg, S., & McGlone, M. S. (2001). *Understanding figurative language: From metaphor to idioms*. Oxford psychology series, 36. New York: Oxford University Press.

Gramley, S., & Pátzold, M. (2003). *A survey of modern English*. New York, London: Routledge.

Grant, L. E. (2007). In a manner of speaking: Assessing frequent spoken figurative idioms to assist ESL/EFL teachers. *System*, *35*(2), 169-181.

Gries, S. T. (2009). What is corpus linguistics? *Language and linguistics compass*, *3*(5), 1225-1241.

Lakoff, G. & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

Lewis, M. (1993). *The lexical approach*. UK: Language teaching publications.

Lim, E. A. C., Ang, S. H., Lee, Y. H., & Leong, S. M. (2009). Processing idioms in advertising discourse: Effects of familiarity, literality, and compositionality on consumer ad response. *Journal of Pragmatics*, *41*(9), 1778-1793.

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 & London: Routledge.

Maisa, S., & Karunakaran, T. (2013). Idioms and importance of teaching idioms to ESL students: A study on teacher beliefs. *Asian Journal of Humanities and Social Sciences (AJHSS)*, *1*(1), 110-122.

Moon, R. (1998). *Fixed expressions and idioms in English: A corpus-based approach*. New York: Oxford University Press.

Ovando, C. J. & Collier, V. P. (1985). *Bilingual and ESL classrooms*. New York: McGraw-Hill Book Company.

Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge: Cambridge University Press.

Schmitt, N. (2004). *Formulaic sequences: Acquisition, processing, and use*, *9*. Amsterdam, Philadelphia: John Benjamins.

Scott, M. (2012). *WordSmith Tools (Version 5.0)* [Computer Software]. Available from http://www.lexically.net/software/index.htm.

Simpson, R. C., Briggs, S. L., Ovens, J., & Swales, J. M. (2002). Michigan corpus of academic spoken English [Data file]. Available from University of Michigan Website, 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), 419-441.

Spears, R. A. (1993). *NTC's dictionary of phrasal verbs and other idiomatic verbal phrases*. Illinois, US: National Textbook Company.

Spears, R. A. (2000). *NTC's dictionary of American slang and colloquial expressions*. US: NTC Publication Group.

Teodorescu, A. (2015). Mobile learning and its impact on business English learning. *Procedia-Social and Behavioral Sciences*, *180*, 1535-1540.

Thyab, R. A. (2016). The Necessity of idiomatic expressions to English Language learners. *International Journal of English and Literature*, *7*(7), 106-111.

Weisser, M. (2009). *Essential programming for linguistics*. Edinburgh: Edinburgh University Press.

Wood, D. (2002). Formulaic language in acquisition and production: Implications for teaching. *TESL Canada Journal*, *20*(1), 1-15.

Wray, A. (2000). Formulaic sequences in second language teaching: Principle and practice. *Applied Linguistics*, *21*(4), 463-489.

Appendices

Appendix A. *The 50 most frequently used idioms in the spoken genre of COCA with their frequency per million words*

Idiom	Frequency (per million)	Idiom	Frequency (per million)
every last (or single)	33.61	do away with something	5.38
(all) of a sudden	31.17	go public	5.13
on (or off) the air	28.25	first off	5.04
the bottom line	20.14	look someone in the eye (or face)	4.89
big deal	19.49	speak for oneself	4.77
at the end of the day	16.71	the old days	4.66
on (the) record	15.79	in (or out of) perspective	4.56
last thing	11.40	behind closed doors	4.55
behind the scenes	10.85	get the message	4.46
in the long run (or term)	10.85	from day one	4.26
be history	9.98	come (or spring) to mind	4.20
and all that (or and that)	9.39	for the record	4.10
all over the place	9.02	a heck of a —	4.01
on someone's mind	7.88	give someone a break	3.99
if anything	7.56	on the side	3.96
once (or every once) in a while	7.49	a (or one) hell of a —	3.70
by and large	6.71	off the hook	3.67
the American dream	6.53	to say the least (or the least of it)	3.63
going on — (or going on for—)	6.13	on (or from) the sidelines	3.60
close to (or close on)	6.12	the real thing	3.58
let it go (or pass)	5.98	for free	3.45
under fire	5.91	— as we know it	3.42
for real	5.91	go (to) it	3.41
in the short run (or term)	5.76	first up	3.36
be good news	5.40	round (or around) the clock	3.34

Appendix B. *The 50 most frequently used idioms in the fiction genre of COCA with their frequency per million*

Idiom	Frequency (per million)	Idiom	Frequency (per million)
big deal	13.18	look someone up and down	5.27
(all) of a sudden	12.18	be history	5.14
every last (or single)	12.11	all over the place	5.10
meet someone's eye (or eyes or gaze)	11.98	have something on one's mind	5.08
look someone in the eye (or face)	11.98	on (or off) the air	5.05
on someone's mind	11.26	lose one's mind (or one's marbles)4.90	4.98
once (or every once) in a while	10.39	stay put	4.90
let it go (or pass)	10.10	set (or put) one's hand to	4.90
a (or one) hell of a —	9.60	come (or spring) to mind	4.83
out of one's mind	9.20	not give a damn (or hoot)	4.75
open the door to	8.25	in (or into) the open	4.67
be there for someone	8.25	the real thing	4.66
keep (or lose) one's head	8.15	if anything	4.62
make (or pull) a face (or faces)	7.56	and all that (or and that)	4.07
on the side	7.44	pissed off	4.01
do someone a favor	7.12	close to (or close on)	3.95
the old days	7.04	for real	3.94
on the horizon	6.87	clap (or lay or set) eyes on	3.93
know better than to	6.65	for free	3.86
for all someone knows	6.64	mean to say	3.85
before you know where you are (or before you know it)	5.96	get the hell out (of)	3.83
snap (or click) one's fingers	5.94	the likes of	3.79
take someone aback	5.78	give (or lend) a hand	3.79
take it easy	5.48	kick someone's ass (or butt)	3.74
off (or out of) one's head	5.30	hold (or keep) someone or something at bay	3.67

Appendix C. *The 50 most frequently used idioms in the magazine genre of COCA with their frequency per million*

Idiom	Frequency (per million)	Idiom	Frequency (per million)
the bottom line	11.05	stay put	4.22
every last (or single)	10.40	do the trick	4.20
in the long run (or term)	9.30	behind the scenes	4.16
big deal	8.68	in the short run (or term)	4.10
the likes of	7.25	hold (or keep) someone or something at bay	3.75
for free	6.91	pave the way for	3.66
be there for someone	6.85	keep (or lose) one's head	6.63
(all) of a sudden	6.71	rule of thumb	3.61
from scratch	6.23	hold one's own	3.55
at the end of the day	6.13	follow suit	3.55
last thing	6.05	to boot	3.49
close to (or close on)	5.94	nothing less than	3.46
once (or every once) in a while	5.60	the old days	3.44
on someone's mind	5.49	look someone in the eye (or face)	3.42
if anything	5.47	in (or out of) sync	3.33
go public	5.38	be good news	3.29
on (the) record	5.19	do away with something	3.27
on the horizon	5.14	come of age	3.25
on the side	5.12	think twice	3.24
fall short (of)	4.94	in (or out of) perspective	3.19
come (or spring) to mind	4.75	the hard way	3.15
one's best bet	4.72	— as we know it	3.05
on (or off) the air	4.52	across the board	3.01
the real thing	4.50	all over the place	3.00
leave (or make) its (or one's or a) mark	4.24	under fire	3.00

Appendix D. *The 50 most frequently used idioms in the newspaper genre of COCA with their frequency per million*

Idiom	Frequency (per million)	Idiom	Frequency (per million)
the bottom line	12.30	leave (or make) its (or one's or a) mark	3.94
every last (or single)	9.84	set the stage	3.94
big deal	9.57	pave the way for	3.82
on (or off) the air	8.83	the old days	3.71
in the long run (or term)	8.70	do away with something	3.57
(all) of a sudden	8.68	the real thing	3.52
the likes of	7.83	across the board	3.49
for free	7.58	follow suit	3.45
go public	7.57	come of age	3.43
fall short (of)	7.42	make (or miss) the cut	3.42
from scratch	6.41	in (or out of) perspective	3.36
at the end of the day	6.34	make (both) ends meet	3.32
on (the) record	6.30	come (or spring) to mind	3.26
if anything	6.09	for the record	3.17
close to (or close on)	6.05	raise one's eyebrows (or an eyebrow)	3.14
behind the scenes	5.63	all over the place	3.11
last thing	5.47	by and large	3.09
under fire	5.37	in (or out of) key	3.00
on someone's mind	5.00	think twice	2.85
on the side	4.84	be good news	2.83
in the short run (or term)	4.62	hold one's own	2.77
on (or from) the sidelines	4.54	round (or around) the clock	2.72
on the horizon	4.29	catch fire	2.71
once (or every once) in a while	4.05	open the door to	2.65
the American dream	4.05	behind closed doors	2.61

Appendix E. *The 50 most frequently used idioms in the academic genre of COCA with their frequency per million*

Idiom	Frequency (per million)	Idiom	Frequency (per million)
in the long run (or term)	9.56	on the horizon	2.39
in (the) light of	7.38	by the same token	2.34
in the short run (or term)	4.54	in (or out of) key	2.32
as it were	5.51	to name (but) a few	2.31
fall short (of)	4.99	take root	2.25
set the stage	4.79	follow suit	2.16
close to (or close on)	4.72	the American dream	2.16
if anything	4.62	from scratch	2.09
pave the way for	4.10	trial and error	2.04
by and large	4.08	be there for someone	2.03
for free	3.60	on (the) record	1.92
come (or spring) to mind	3.52	on someone's mind	1.90
do someone or something justice (or do justice to someone or something)	3.41	once and for all (or once for all)	1.83
the bottom line	3.29	the pros and cons	1.77
under the auspices of	3.24	it remains to be seen	1.69
in tandem	3.23	behind the scenes	1.67
do away with something	3.12	— as we know it	1.65
nothing less than	3.05	hearts and minds	1.65
in force	2.74	at a (or the) crossroads	1.63
one and the same	2.67	at the end of the day	1.63
in (or out of) perspective	2.63	state of the art	1.48
come of age	2.58	come (or get) to grips with	1.46
in the final analysis	2.49	the likes of	1.42
open the door to	2.46	across the board	1.39
every last (or single)	2.39	leave (or make) its (or one's or a) mark	1.37