Rafatbakhsh and Ahmadi *Asian-Pacific Journal of Second and Foreign Language Education* (2019) 4:11
https://doi.org/10.1186/s40862-019-0076-4

Asian-Pacific Journal of Second and Foreign Language Education

ORIGINAL ARTICLE Open Access

A thematic corpus-based study of idioms in the Corpus of Contemporary American English

Elaheh Rafatbakhsh and Alireza Ahmadi*

* Correspondence: arahmadi@shirazu.ac.ir
Department of Foreign Languages and Linguistics, Faculty of Literature and Humanities, Shiraz University, Shiraz 7194684795, Iran

Abstract

The traditional approach to presenting idioms relies mainly on teachers' or materials writers' judgement, one-by-one and quite incidentally; and the existing teaching materials and references for idioms are mostly intuition-based. However, a more recent approach to better teaching and learning idioms is to present them under categories of their common themes and topics. Corpus linguistics can be of much contribution through helping the design and development of more authentic and systematic materials using comprehensive corpora which are typically the best representatives of the target language. In this connection, the present study aimed at searching for the thematic index of 1506 idioms under 81 categories at the end of the Oxford Dictionary of Idioms in the largest freely available corpus, i.e. the Corpus of Contemporary American English (COCA), composed of more than 520 million words. To this end, we used a manuscript in PHP written by a professional computer programmer especially for this purpose. The findings yielded a list of idioms sorted based on their frequencies under their theme-based categories. To focus on the more frequently-used idioms of various themes in real contexts, materials designers, teachers, and learners of English can benefit from this idiom list in textbooks and classroom activities.

Keywords: Idioms, Corpus of Contemporary American English (COCA), Frequency list, ESL/EFL teaching, Materials development

Introduction

An idiom is defined as a "constituent or series of constituents for which the semantic interpretation is not a compositional function of the formatives of which it is composed" (Fraser, 1970; p.22). Idiomaticity is a topic of some research and studies in the areas of linguistics, psycholinguistics, developmental psychology, and neuropsychology (Cacciari, 1993). "The importance of idioms in any language cannot be doubted. Their ubiquity makes them anything but a marginal phenomenon, and surely a linguistic theory has an obligation to explain them in a natural way" (Chafe, 1968; p.111).

It is believed that the more words a learner knows, the more proficient they are in the process of language learning (Laufer & Goldstein, 2004; Nation & Meara, 2002). This assumption does not take the different combinatory possibilities of words into consideration (Daskalovska, 2011). A successful language learning mastery includes a crucial component of learning formulaic sequences such as idioms, collocations, and compounds (Wray, 2000).

Idioms provide the language with variety and imagination (Cooper, 1999) and their learning embodies learning a culture (Crystal, 1997; Glucksberg & McGlone, 2001; Ovando & Collier, 1985). Learning idioms can lead to native-like proficiency since idioms are learned in chunks, and consequently, they are retrieved from the learners' memory without hesitations which can lead to fluency (Boers, Eyckmans, Kappel, Stengers, & Demecheleer, 2006; Cain, Towse, & Knight, 2009; Lim, Ang, Lee, & Leong, 2009; Teodorescu, 2015). On the other hand, when language learners learn the words individually, they may need more time to retrieve them and make proper sentences. Translators or language teachers cannot afford to ignore idioms or idiomaticity if their aim is natural use of language (Fernando, 1996).

Furthermore, idioms are considered to be a particular kind of lexical items. Lexical items are "socially sanctioned independent units". Language is not individually defined; it is rather a social phenomenon. Lewis (1993) emphasized the importance of "lexical approach" in language teaching and learning. According to this approach, although language can be sub-divided into sentences, turns, morphemes and phonemes, "lexical items" can be considered as the minimal units for certain syntactic purposes. Larger sequences are too large for analyses and shorter ones are too short. The multiword units can be analyzed but researchers (Lewis, 1993; Nattinger, 1988) believe that these units should be perceived as single, unanalyzed wholes which is in turn quite important in the process of language teaching and learning.

Although figurative expressions may appear quite arbitrary, there exist certain structures and organizations among them and a large number of these expressions have common metaphoric themes (Kovecses & Szabco, 1996). For example, there is a wide range of idioms related to themes of nature, animals, body parts, sports, specific names, food, colors, and all the senses which are used to describe personality, appearance, work, health issues, and many more (O'Dell & McCarthy, 2010). Moreover, idiomatic expressions often indicate and reflect social norms, beliefs, attitudes, and emotions (Glucksberg & McGlone, 2001). There is a taxonomy that suggests classifying teaching approaches into two broad categories of non-semantic (Szczepaniak, 2006; Tran, 2013; Vrbinc & Vrbinc, 2011) and semantic approaches (Boers & Stengers, 2008; Panou, 2014). In the non-semantic approaches, the idioms are provided for the learners one-by-one and quite incidentally through the materials, while in the semantic approaches, the idioms are presented by their shared meanings or themes. Therefore, the semantic approaches, which are based on classification of idioms according to metaphors, their source domains and origins can enhance idiom learning and deeper retention (Boers, 2013).

As a large number of idioms are not intelligible to learners at first sight and their meanings typically cannot be guessed through the analysis of the components, teaching and learning idioms have always been a problematic and challenging part of language. There are several drawbacks and limitations in the teaching textbooks concerning the choice of idioms as well as their presentation which have made their teaching and learning even more difficult (Cameron & Low, 1999; Simpson & Mendis, 2003). Therefore, there is still a lot of room for more research and improvement on idioms in teaching methods and materials, particularly in EFL/ESL contexts. Applied and corpus linguistics can offer solutions to this problem by identifying frequencies and patterns of idiom use in order to give priorities in teaching and learning contexts (Liu, 2003).

Johansson (2009) has summarized the uses of corpora in language acquisition in the following figure (Fig 1).

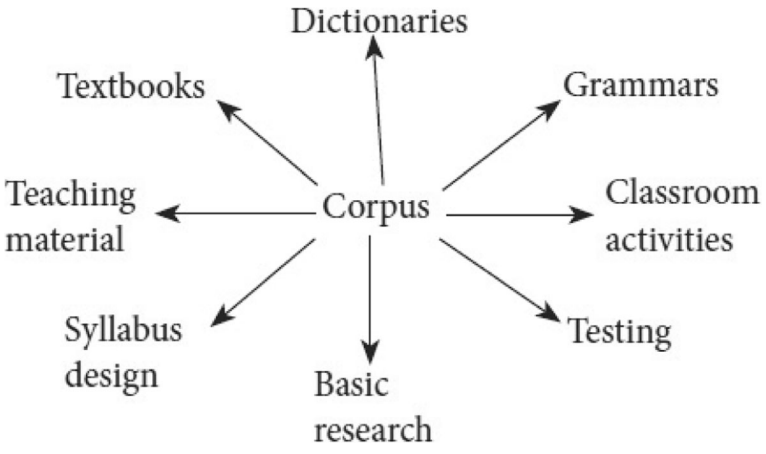

Fig. 1 Johansson: Uses of corpora of relevance for language teaching

He argues that the corpus can be used in almost every aspect of language teaching and acquisition including materials development, testing, and so far.

Corpus search results and patterns can effectively replace teachers' and material developers' intuition when preparing the materials. A teacher can determine the level of importance of the materials for language teaching using corpus linguistics (Biber & Reppen, 2002). In this connection, the present corpus-based study aimed at identifying the frequencies of the idioms' categories in the large Contemporary Corpus of American English (COCA) and reporting the most frequent ones for such purposes.

Literature review

Defining idioms

Scholars have always had difficulty defining what an idiom is. Although there exist a lot of definitions, it is sometimes impossible to differentiate between collocations, phrasal verbs and idioms. In her dissertation, Grant (2003), after summarizing all the definitions of an idiom, came a conclusion that linguists have not reached a consensus on idiom definition and classification for language teachers and learners.

Fernando (1996) considers idioms as "conventionalized multi-word expressions often, but not always non-literal" (p. 1). Gramley and Pátzold (2003) have added more to that definition and define 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). Semantically speaking, scholars have proposed different scales or continuum of idiomaticity (Alexander, 1987; Cowie, Mackin, & McCaig, 1983; Fernando, 1996; Moon, 1998b; Wood, 1981). Grant (2003) summarized the scales used by such scholars into six categories of a) semi-idioms including at least one word connected to its literal meaning (e.g. white lie), b) semi-opaque idioms whose meanings can be guessed but not easily (e.g. sail too close to the wind), c) pseudo idioms including an element that has no meaning on its own (e.g. spic and span), d) pure idioms, well-formed idioms, or idioms of decoding that have both literal and non-literal meaning (e.g. kick the bucket), e) full idioms which consist of constituents whose ordinary meanings are not related to the idioms' semantic

interpretations (e.g. butter up) and f) figurative idioms that have figurative meanings besides current literal interpretations (e.g. catch fire).

There are three themes repeated in all the definitions of idioms, compositionality, institutionalisation, and degree of frozenness or fixedness. First, the idioms are non-compositional since their meanings are not the sum of the meanings of their parts; secondly, they are institutionalised which means they are commonly used by a large number of people in a speech community; and finally, the idioms are frozen and fixed i.e. but the degree of their frozenness varies (Grant, 2003).

There is a blur line between idioms and other multiword units and since the mentioned characteristics might be common among them, it is difficult to differentiate between idioms and multiword units. This lack of unanimous agreement among scholars causes a lot of difficulties for teachers and learners on what an idiom is. Therefore, their only criterion for the choice of an idiom is dictionaries. However, the great number of existing idioms in the dictionaries is quite overwhelming and the attempts to prioritize them based on the categories and frequencies can be of great help in this respect.

Teaching and learning idioms

As mentioned earlier, formulaic language such as idioms should be taught in EFL contexts in order to increase the fluency and naturalness of leaners' language. Teaching and learning idioms have a lot of advantages for learners such as promoting communicative competence, proficiency, fluency and familiarizing them with the target language culture (Bardovi-Harlig, 2002; Fernando, 1996; Liu, 2008; Moon, 1998a; Schmitt, 2004; Thyab, 2016; Wood, 2002; Wray, 2000). Teaching and learning idioms are important and significant, since not only can lack of knowledge of idioms cause serious comprehension problems and misunderstandings in many contexts, even rich in clues, but also the use of idioms and figurative idioms in particular, is not as infrequent as it has been assumed (Boers, 2013). The results of an exploratory research (Maisa & Karunakaran, 2013) on the importance of teaching idioms to ESL students based on teachers' perspectives showed that teachers believed teaching idioms to undergraduate students as an integral part of vocabulary teaching, lead to more fluent speaking and writing. As a result, they suggested the inclusion of idioms in dialogues, readings, and stories in the curriculum.

Boers (2000) carried out three experiments on figurative expressions in EFL contexts and all three studies involved two parallel groups of control and experimental. The participants of three studies were each 118, 73, and 74 intermediate learners of English respectively. In the first experiment, both groups were asked to read a text about emotions. Following that, the experimental group were given a list of expressions and vocabulary from the text, organized by their metaphoric themes while the control group received the same vocabulary but not in an organized way. The participants studied and discussed the expressions and subsequently took a cloze test based on the list of vocabularies and expressions. In the second experiment, the two groups received a list of vocabulary all on economic trends, however, unlike the control group, the experimental group had their list sorted based on the expressions' source domains. To test the effectiveness of the experiment, they were asked to write an essay describing some graphs and the use of the mentioned expressions were examined and compared accordingly. Finally, in the third experiment, pupils were presented with a set of multiword verbs (prepositional and

phrasal verbs). The experimental group received a list categorized by the headings of underlying orientational metaphors while the control group had the same list alphabetically. Following that, they all took a cloze test on the same topic. The result of all the three experiments indicated superior retention of the figurative expressions in experimental groups who received the lists sorted based on underlying metaphoric themes. That is because determining source domains lead to deep-level cognitive processing which in turn enhanced memory storage and learning.

Corpus-based studies on idioms

Although searching for idioms in corpora is a difficult and complex process due to the complex nature of the idioms, some studies have focused on searching the frequencies of idioms and their patterns of use in various corpora. For instance, Baddorf and Evens (1998) selected three corpora of 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) and searched 30 idioms and their syntactic variants in these corpora using a computer program written for this purpose. They reported the idioms and their frequencies in details from the most to the least frequent.

Another main corpus-based study was carried out by Moon (1998b) 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, and 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, more revisions of existing models and descriptions should be made and the importance of the roles FEIs play in discourse should not be underrated.

Later, Liu (2003) in a study to identify pure semiliteral and literal idioms consulted four major contemporary English idiom dictionaries and three English phrasal verb dictionaries. Subsequently, he searched for such idioms in three contemporary spoken American English corpora, Corpus of Spoken, Professional American English, Michigan Corpus of Academic Spoken English, and Spoken American Media English. The findings of the study suggested four lists of the most frequently used idioms and their patterns of use followed by recommendations for improving quality of teaching, materials and references in terms of idiom selection, meaning, explanation, and the examples. The researcher criticized teaching and including idioms in materials based on mere intuition.

In another corpus study of idioms by Simpson and Mendis (2003), the researchers selected the idioms based on three factors of compositeness or fixedness, institutionalization and semantic opacity. The corpus used was the Michigan Corpus of Academic Spoken English (MICASE) consisting of 1.7 million words of academic discourse. The result suggested two idiom lists: The idioms particularly useful for English for academic purposes curricula and the idioms occurring four or more times in MICASE. Based on their findings, they recommended the use of corpus studies in teaching and learning idioms.

In a major corpus-based search of idioms, to prioritize teaching and learning idioms for ESL/EFL teachers and learners based on their frequencies of use in a corpus by

Grant (2007), three groups of core idioms (non-compositional, non-figurative), figuratives (non-compositional, figurative), and ONCEs (one non-compositional element, may also be figurative) were used. The corpus for this study was the spoken genre of the British National Corpus (BNC) and the results were then compared to Liu's (2003) and Simpson and Mendis's (2003) lists. 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.

Additionally, the number of color idioms were studies and also compared in the Corpus of Contemporary American English, British National Corpus, and TIME Magazine Corpus of American English by Vaclavikova (2010). The results manifested higher popularity of idioms in American English than in British English. Furthermore, the overall number of color idioms was greater in newspapers and magazines than in academic texts.

It can be very demanding and time consuming to search for a large number of idioms in large corpora. To date, not enough corpus-based studies have been conducted on idioms and the existing research has been limited to only some specific types of idioms such as core idioms or figuratives. Moreover, the corpora chosen for the studies have been restrained to the academic or spoken sections which are not very large in size. Therefore, the purpose of the present study was to uncover the frequencies of all the idioms of the thematic index of the Oxford Dictionary of Idioms, grouped based on their topics and themes, along with their variations and forms in the large Corpus of Contemporary American English (COCA). The search was done via a manuscript written by a computer programmer for this purpose. The search result is a list of 1506 idioms in 81 theme-based categories along with their frequencies per million, sorted from the most frequent to the least in each category.

Method

This section begins by introducing the corpus and the idiom dictionary used in the present study. It then provides information on how the idioms were prepared and searched for in the corpus.

The corpus

The research was based on the Corpus of Contemporary American English (COCA) including 20 million words each year from 1990 to 2017; however, the purchased version of COCA for this study was composed of 520 million words in 220,225 texts from 1990 to 2015. COCA was created by Mark Davies, Professor of Corpus Linguistics at Brigham Young University. Currently, COCA is the most recent, comprehensive and balanced corpus of English language that exists. This corpus is divided evenly between 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 conversation from more than 150 different TV and radio programs (examples: All Things Considered (NPR), Newshour (PBS), Good Morning America (ABC), Today Show (NBC), 60 Minutes (CBS), Hannity and Colmes (Fox), Jerry Springer, etc.). 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, first chapters of first edition books 1990-present, and movie scripts. 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. Time, Men's Health, Good Housekeeping, Cosmopolitan, Fortune, Christian Century, Sports Illustrated, etc. are some examples of magazines used for this purpose. For newspapers, 106 million words (105,963,844), a good mix of various sections such as local news, opinion, sports, financial, etc. of 10 newspapers including USA Today, New York Times, Atlanta Journal Constitution, San Francisco Chronicle, etc. is included. 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 (e.g. a certain percentage from B (philosophy, psychology, religion), D (world history), K (education), T (technology), etc.), both overall and by number of words per year. 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. Table 1 summarizes COCA, its genres, and the number of words in each from 1990 to 2015.

Idiom dictionary

As the main source of the idioms, the Oxford Dictionary of Idioms, 2nd ed., was chosen since it was the latest dictionary of idioms available to the researchers in digital format. This dictionary includes about 5000 British and American idioms along with their definitions, explanations and some with illustrative quotations and examples. Background information of some of the idioms is also attached in a box. Overall, a comprehensive picture of the idioms is presented in this dictionary. Moreover, at the end of the book, 1506 idioms with the same topics and themes are grouped together under 81 categories. This gives the users a vivid picture of those areas and aspects of life that have generated a rich variety of figurative expressions. The target idioms of the current corpus-based study were all the 1506 idioms included in the index. The 81 themes and the number of the idioms in each are illustrated in Table 2.

The number of idioms in each theme differs ranging from 7 to 39 with a mean of 28. The theme of 'pregnancy' includes the least number and the 'misfortune and adversity' the largest number of idioms in this index.

Data collection procedure

Obtaining the frequencies of the idioms involved three main steps. First, a computer program was written; then the idioms were manually prepared to be searched by the designed system; and finally, they were searched for in the whole corpus of COCA.

Idioms transformation and coding procedure

First, all idioms of the index were extracted from the digital version of the Oxford Dictionary of Idioms manually and a list of 1506 idioms was created. Each entry was checked again in the dictionary to add other variations of the idiom if present. For instance, the idiom "have seen better days" in the index has also another variation "have

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

Table 2 Themes of Oxford Dictionary of Idioms index

Theme	Idioms
Action	28
Age	20
Ambition	12
Anger and annoyance	37
Anxiety and worry	20
Appearance	23
Argument and conflict	32
Beauty	12
Boastfulness and conceit	17
Bribery, corruption, and extortion	19
Caution	18
Certainty	17
Change	24
Chaos and disorder	25
Class	13
Clothes	11
Cooperation	26
Courage	18
Crime and punishment	20
Crisis	19
Critics and criticism	24
Danger	25
Death	29
Debt	11
Deception and lying	25
Doubt and uncertainty	10
Duty and responsibility	13
Embarrassment, shame, and humiliation	18
Equality	15
Excess and extravagance	21
Expense	19
Experience	17
Family	18
Fate and chance	8
Food	15
Fools and foolishness	19
Foresight and the future	7
Forgiveness and reconciliation	10
Futility	30
Gossip and rumor	8
Happiness, pleasure, and enjoyment	33
Haste and speed	32
Health and illness	15
Honesty	14

Table 2 Themes of Oxford Dictionary of Idioms index *(Continued)*

Theme	Idioms
Hope and optimism	14
Hypocrisy	8
Indecision and prevarication	13
Intelligence and knowledge	16
Jealousy and envy	7
Justice	10
Language, speech, and conversation	22
Laziness	9
Love	14
Madness	19
Marriage	8
Misfortune and adversity	39
Mistakes	18
Money, wealth, and prosperity	27
Opportunity	29
Poverty	15
Power	28
Pregnancy	7
Preparation and readiness	26
Reputation and fame	15
Revenge and retribution	16
Secrecy	29
Self-interest	25
Strength	18
Success	30
Surprise	13
Thoroughness	14
Time	19
Traitors and treachery	17
Travel and transport	24
Unhappiness and disappointment	17
Violence	18
Warfare	12
Weakness	16
Weather	9
Work and employment	27
Youth	11

known better days". Therefore, all the idioms of the list were updated based on their entries in the dictionary.

Normally, concordancers are used to extract frequencies from a given corpus. Concordancers are computer programs for text analysis usually used in corpus linguistics to retrieve alphabetically or otherwise sorted lists of linguistic data from a corpus. So far, four

generations of concordancers have been developed; however, the third generation tools such as *WordSmith Tools* (Scott, 2012) and *AntConc* (Anthony, 2012) have some limitations. For instance, they are not able to manage large corpora of more than 100 million words because of their architecture. The fourth generation tools such as *corpus.byu.edu* (Davies, 2015), *CQPweb* (Hardie, 2013), and *SketchEngine* (Kilgariff, 2013) are mostly web-based and include a single corpus interface with their own specific controls and operation characteristics that cannot be easily adapted to every use. Given the mentioned shortcomings of the concordancers, researchers such as Anthony (2009), Biber, Conrad, and Reppen (1998), Gries (2009), and Weisser, M. (2009) maintain that the best solution for the corpus linguists is to program and develop their own tools for text analysis to fit their own needs.

Since there were a large number of idioms intended to be searched for in the large COCA, a professional computer programmer wrote a script in PHP for this purpose. In order for the search to be thorough and to include all forms and variations of each idiom, all the idioms were manually prepared for the program. The preparations were as follows:

- The main verbs of the idioms that could take different forms depending on the context were identified and capitalized (e.g. the idiom "shake a leg" was rewritten as "SHAKE a leg"), and the system would search for idioms with all the grammatical forms.
- Some idioms included possessive adjectives, subject and object pronouns or words such as something, someone, somebody, one, etc. that were not fixed and could be replaced by other words in the context. These words were changed into an asterisk symbol before the search (e.g. the idiom "rattle someone's cage" were rewritten as "RATTLE * cage"). Different samples of some of these idioms were studied from the corpus and it was decided that the asterisks represent from no words up to three words during the search. 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; or the idiom "be one's own man (or woman or person)" might be "be HIS OWN man" or "be VERY MUCH HIS own man".
- Some idioms may include interchangeable elements with the same overall meaning which can create nodes when searching for the frequencies. In case nodes were present in the idioms, the symbol "|" was used to separate the two or more options and search for them all (e.g. the idiom "cry over spilt (or spilled) milk" changed into "CRY over spilt|spilled milk").
- COCA has its own specific written format; e.g. contractions and possessives such as "'s, n't, 're, etc." are separated by a space. Hence, the idioms including these contractions were rewritten as they are in the corpus (e.g. the idiom "Caesar's wife" was written as "Caesar 's wife").
- In order to cover more related idioms some articles at the beginning of the idioms were omitted when possible (e.g. the idiom "a black sheep" was changed to "black sheep").
- Since some words such as rumor, honor, color, etc. have two different spellings, they were written as nodes separated by '|' symbol in order to be more

comprehensive e.g. the idiom "with flying colours" were rewritten as "with flying colours|colors".)

In addition, a list of all English verbs and their forms (past tense, past participle, third person, and gerund) was added to the process to address the grammatical forms of each idiom. Some verbs such as *have*, *to be*, and modals also had their negatives and contracted forms 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.

As an example, for the idiom "settle (or pay) a (or the) score", the formula was written as "SETTLE|PAY a|the score" and the system searched for all the following idioms in the corpus and provided a frequency of all of them for each year and each genre separately and summed up the figures automatically:

- Settle a score, settles a score, settled a score, settling a score
- Settle the score, settles the score, settled the score, settling the score
- Pay a score, pays a score, paid a score, paying a score
- Pay the score, pays the score, paid the score, paying the score

Searching the corpus First, the accuracy of the written script was tested, comparing some random idioms' frequencies calculated by the current system with the frequencies received from the concordancer of the website of COCA. The matching results manifested that the script worked very well.

Afterwards, the written script broke the big COCA into smaller pieces and then used regular expressions -strings of text that allow creating patterns that help match, locate, and manage text in programming- to search for the frequencies of each idiom in the COCA. The designed system then provided a comprehensive spreadsheet of statistics with the frequencies of the idioms for each year from 1990 to 2015. The result of the search was a table with 1506 rows and 127 columns. Each genre in each year had a cell with a frequency result for each idiom. The frequencies were then summed up and the idioms in each theme category were sorted from the most frequent to the least. Some idioms such as "have a ball" or "past it" have both idiomatic and non-idiomatic meanings in different contexts. Hence, these idioms were manually searched for in the website of COCA and by reading all the existing contexts, their frequencies were recalculated again and their literal and non-idiomatic uses were subtracted. Some instances of both uses in the context are as follows:

1. (Spoken genre: NBC Today, 2011) We're back from Seattle. We had a ball in Seattle. Thank you, everybody, for coming and being with us while we taped two shows.
2. (NEWS genre: Denver Post, 2009) In order for Johnson, Colorado State's star running back, to get a job in the NFL, he would have to forget - at least for now - how to run like a football player. # " Being a running back, you're used to carrying the ball, and when I first got there, I was running the 40 like I had a ball in my hands, " Johnson said. "

In the first context above, the idiom means having fun and enjoying oneself, while in the second the meaning of the phrase is quite literal. Moreover, the frequency of some of the top idioms that included asterisks were checked in the website of COCA again in order to omit irrelevant words that had substituted an asterisk.

A sample search result of an idiom in the corpus is shown below. An idiom such as "rest on one's oars" was repeated 5 times in the corpus. The system searched for all the verb forms of the verb "rest" and all the possibilities of the words that could replace "one's". The five sample sentences are as follows:

1. (Fiction genre: Field & Stream, 1997) He *rested on the oars* and watched the water for the length of time it would have taken to make a couple dozen casts, to search under the alders, along the seam, beyond the chop.
2. (Fiction genre: Arkadians, 1995) By the time Oudeis called a halt, Lucian's muscles were twitching in protest, and he was glad enough to *rest on his oars*.
3. (Fiction genre: Slow Waltz in Cedar Bend, 1993) Just turned in my grades and *resting on my oars*.
4. (NEWS genre: Christian Science Monitor, 1993) But we are in no way guaranteed to keep in that position if we *rest on our oars*.
5. (NEWS genre: San Francisco Chronicle, 1991) My concern is that the industry has not done anything to improve the (accident) situation for 6 years -- they're kind of *resting on their oars*, "said William Pugh, former top railroad investigator at the National Transportation Safety Board and now a private consultant.

Findings

By searching all the forms and variations of each idiom, the researchers attempted to reveal a comprehensive picture of the frequencies of the index. However, since idioms have a specific unpredictable nature at times, the full cover cannot be claimed.

To show the relative proportion of the idioms in the whole corpus of COCA, the frequencies of all the idioms were calculated per million and inserted in a table. Due to the very large size of the table, only the results of the first three themes, action, age, and ambition are depicted in Table 5 in Appendix.

Based on the search results, most of the idioms of the Oxford Dictionary index have the frequency of less than 1 per million in the whole corpus except for 17 idioms. Table 3 reports the idioms that had a frequency of above 1 per million in COCA along with their themes and frequencies.

The idiom "behind the scenes", belonging to the theme of secrecy, seems to be the most frequent of all the idioms of the index with the frequency of 4.71 per million. Two instances of the idiom together with their sources, genres and years are given here.

- (NEWS genre: USA Today, 2015) Everyone looks flawlessly glamorous for the Academy Awards, but behind the scenes, millions of dollars and thousands of hands are making sure the show goes off without a hitch.

Table 3 The idioms with the frequency of more than 1 per million

Idiom	Theme	Frequency (per million)
Behind the scenes	Secrecy	4.71
Under fire	Critics and criticism	3.22
Set the stage for	Preparation and readiness	2.86
Out of your mind	Madness	2.77
Over the top	Excess and extravagance	2.72
Think twice	Caution	2.33
Round (or around) the clock	Time	2.13
Behind closed doors	Secrecy	2.03
Hit the road	Travel and transport	1.56
Play ball	Cooperation	1.38
Call the shots (or tune)	Power	1.21
On the same page	Cooperation	1.18
In a nutshell	Language, speech, and conversation	1.18
A slippery slope	Misfortune and adversity	1.14
Turn the corner	Crisis	1.12
On a roll	Success	1.10

- (Magazine genre: People, 2012) Everyone saw me on TV or read articles, and it was all about my great marriage, the white picket fence, all this success and my perfect life. But behind the scenes, it was a struggle, "says Vonn, now 28, over lunch.

The next most frequent idiom in the list is "under fire", related to theme of criticism, which was repeated 3.22 per million in the whole corpus.

- (News: Atlanta Journal Constitution, 2009) The Bush administration came under fire in 2006 and 2007 for what appeared to be the politically motivated firings of several U.S. attorneys, amid accusations that administration officials were trying to turn the prosecutors into partisan agents.
- (Academic: Professional School Counselling, 2005) To be sure, middle school counselors' efforts are most accepted and become more valuable when harmonious to the focus on academic achievement. This is particularly true as middle schools have come under fire for the declines in achievement during the middle school years.

On the other hand, 234 idioms of the index, for example, "go down a storm" and "knight of the road", appeared to be non-existent in the whole corpus of COCA. Of the 1506 idioms, 15.5% had the frequency of zero, 726 of them (48.2%) occurred between one and 40 times, and 546 idioms (36.3%) repeated more than 40 times in the whole corpus.

The overall result of the corpus search of the themes is presented in Table 4. The themes are sorted based on their overall frequencies in the corpus and the figures show the frequencies per million.

Table 4 The overall frequencies of the themes

Theme	Overall freq.
Secrecy	10.26
Cooperation	7.64
Preparation and readiness	7.26
Boastfulness and conceit	6.30
Crisis	5.76
Success	5.63
Power	5.31
Excess and extravagance	5.11
Change	5.08
Critics and criticism	5.07
Misfortune and adversity	4.95
Strength	4.94
Caution	4.86
Action	4.20
Travel and transport	4.20
Danger	4.09
Madness	3.84
Anger and annoyance	3.75
Time	3.68
Expense	3.64
Hope and optimism	3.58
Happiness, pleasure, and enjoyment	3.29
Haste and speed	3.24
Ambition	3.02
Chaos and disorder	2.95
Argument and conflict	2.91
Death	2.74
Certainty	2.55
Experience	2.48
Class	2.36
Thoroughness	2.29
Appearance	2.25
Honesty	2.18
Language, speech, and conversation	2.16
Work and employment	2.13
Duty and responsibility	2.07
Embarrassment, shame, and humiliation	2.06
Self-interest	1.99
Equality	1.97
Opportunity	1.91
Mistakes	1.89
Violence	1.89
Deception and lying	1.82
Family	1.8

Table 4 The overall frequencies of the themes *(Continued)*

Theme	Overall freq.
Revenge and retribution	1.76
Bribery, corruption, and extortion	1.71
Money, wealth, and prosperity	1.71
Futility	1.67
Surprise	1.64
Courage	1.59
Reputation and fame	1.58
Crime and punishment	1.53
Marriage	1.51
Pregnancy	1.47
Unhappiness and disappointment	1.32
Age	1.28
Debt	1.27
Health and illness	1.23
Weakness	1.12
Forgiveness and reconciliation	1.1
Indecision and prevarication	1.09
Laziness	1.09
Love	1.01
Foresight and the future	0.97
Gossip and rumor	0.96
Anxiety and worry	0.93
Justice	0.92
Traitors and treachery	0.90
Jealousy and envy	0.89
Intelligence and knowledge	0.87
Poverty	0.85
Food	0.79
Hypocrisy	0.78
Weather	0.77
Warfare	0.76
Youth	0.74
Beauty	0.62
Doubt and uncertainty	0.57
Clothes	0.27
Fate and chance	0.25
Fools and foolishness	0.16

The theme of secrecy, consisting of 29 idioms with the overall frequency of 10.26 per million, is the most frequent theme, while the topic of fools and foolishness including 19 idioms with the overall frequency of 0.16 per million is the least frequent one.

Discussion and conclusions

Searching for the idioms in the thematic index of the Oxford Dictionary of Idioms and their forms and variations in the largest freely-available corpus of English, COCA, led to a frequency list of idioms organized based on 81 topics and sorted by the frequencies of occurrence (Table 5 in Appendix). Overall, among the 1506 idioms searched for in the corpus, 17 had frequencies above 1 per million and 234 idioms were not found at all.

The quantitative comparisons of the findings, idiom lists and the frequencies of the present study with the previous similar ones (Grant, 2007; Liu, 2003; Moon, 1998b; Simpson & Mendis, 2003) are difficult to assess because of the following reasons: The idioms chosen in this research were not similar to the previous corpus search. The criteria of choosing an idiom, the type and number of idioms were quite different in each study. We searched for all types of idioms recorded in the thematic list of the Oxford Dictionary of Idiom, whereas former studies focused on different sets of idioms such as core or figurative idioms. What is more, the corpora (e.g. BNC, MICASE) and the sections or genres (e.g. academic or spoken English) in other studies differed from the current one which makes comparison even more difficult due to the relative infrequency of the idioms.

The present corpus-based study has shed light on the more frequent idioms of various themes in order to ease their teaching and learning. As mentioned earlier, native-like idiomaticity has been the target of language learning in many cases, so collocations, formulaicity and idiomaticity should also be the focus of educators. Prefabricated chunks such as idioms aid learners to improve fluency specifically in spoken language since they are retrieved from the memory with less hesitations (Boers et al., 2006; Mauranen, 2004; Teodorescu, 2015). Based on the fundamental attitudes of the lexical approach the use of awareness-raising activities directing students' attention to the chunks in a given text is essential. When it comes to natural use of language, idioms without a doubt cannot be neglected. Therefore, the result of the present study can benefit English teachers, learners, as well as materials developers in several ways.

First, instead of using intuition to develop the language learning materials, the real use of language from the corpus can provide more authentic sources. To highlight the importance of corpus studies, Liu (2003), after analysis of some teaching and reference materials, criticized the way the idioms were selected which was quite inconsistent since they included less frequent and more transparent idioms but missed some highly frequent or highly opaque items. The selection of materials based on corpus studies would be more systematic and rigorous comparing to the intuition-based sources. It can help in "selecting and sequencing linguistic content, as well as determining relative emphases" (Tsui, 2004, p. 40). The present study produced frequency lists for a large number of idioms (1506 idioms) from a wide variety of topics (81 thematical categories). This can provide language teachers and learners with diverse opportunities to practice idioms and help them prioritize what is more important and beneficial to teach and learn first. Given the short amount of time and the large amount of materials teachers and learners have to deal with, it is quite crucial to first focus on the more useful and frequently repeated parts of the language rather than spending time on the less frequent ones.

Based on the lists provided in this study, language teachers and learners can easily select idioms from a specific theme of their interest knowing how frequent that idiom is in the authentic texts produced by native speakers. This can be of great help in learning idioms which are of more use in different contexts. As Muller-Hartmann and Schocker-von Ditfurth (2004) argue "teachers need to be able to present language as naturalistic examples of the target language, to expose learners to examples of language currently in use, with features which are characteristic of authentic discourse in the target language (p. 28)".

Second, the findings of this study can be of paramount importance in the contexts of EFL where there is often not enough exposure to target language for learners. In such contexts, education should focus more on frequent and authentic parts of the language. This can assist them with learning idioms that are probably more useful to them. Furthermore, information about idiom distribution and frequency "may help students develop a more complete grasp of the idioms or decide to what extent they want to learn and use those idioms" (Liu, 2003, p. 687).

Third, the frequency lists of idioms can be employed by materials developers and language teachers to develop different exercises for learners at different proficiency levels. As argued by Simpson and Mendis (2003), when it comes to pedagogical materials, pupils have proved to respond well to multiple-choice exercises using the idioms' definitions. They also respond well to the items with extracts from the corpus as the stem in which they have to guess the meaning of the idioms. However, the importance of rich contextual clues in the selected sentences from the corpus should be taken into account.

Fourth, test developers can also benefit from the results of this study as they can use the frequency lists in preparing test items. Depending on the proficiency levels of the testes, idioms of different frequencies can be included in the test. Furthermore, different themes may also be considered in preparing test items.

Finally, learners' attention should be drawn to formulaic language during their process of learning and one of the most practical ways to do it is to present the idioms in a theme-based manner. Learning figurative expressions such as idioms by using their common themes can create the possibility of better retention since these topics and themes present a framework and organization for the random lists and make it much easier for the learners to grasp them more deeply (Boers, 2000; Ellis, 1994). Raising the students' awareness for categorizing the idioms based on their topics and themes during the classroom activities, not only can help them learn the idioms more deeply and easily, but also can assist them to be more independent and successful outside the classrooms.

Limitations

The inaccessibility of more up-to-date and recent dictionaries of idioms in digital format was a limitation that needs to be considered in using or generalizing the results of this study. The main source of the idioms of the study was the Oxford Dictionary of Idioms with about 5000 British and American idioms, and the research was based on the Corpus of Contemporary American English (COCA). It is possible that British idioms are not fully and representatively presented in COCA.

Appendix

Table 5 Thematic index of Oxford Dictionary of Idioms along with the frequencies per million

Idiom	Frequency	Frequency per million
Action		
Roll up your sleeves	480	0.91
Take the plunge	367	0.69
Press the button	342	0.65
Put your money where your mouth is	190	0.36
Hit the ground running	184	0.35
Rest on your laurels	169	0.32
Watch someone's smoke	128	0.24
Get cracking	54	0.10
Start the ball rolling	52	0.10
Lead from the front	48	0.09
Keep your nose to the grindstone	40	0.08
Hot to trot	32	0.06
Shake a leg	29	0.05
Strike while the iron is hot	19	0.04
Set the wheels in motion	18	0.03
Put your shoulder to the wheel	16	0.03
Get the show on the road	15	0.03
Hammer and tongs	12	0.02
Get (or pull) your finger out	9	0.02
Rest on your oars	5	0.01
Stir your stumps	4	0.01
Go for the doctor	4	0.01
Have many irons in the fire	1	0
Get weaving	1	0
At the coalface	1	0
No peace for the wicked	0	0
Put (or set) your hand to the plough	0	0
Get the bit between your teeth	0	0
Age		
Over the hill	212	0.40
Have seen (or known) better days	167	0.32
On your last legs	95	0.18
Long in the tooth	63	0.12
Ancient (or old) as the hills	35	0.07
Second childhood	28	0.05
You can't teach an old dog new tricks	18	0.03
Past it	10	0.01
There's no fool like an old fool	10	0.02
Have one foot in the grave	10	0.02
Out of the ark	7	0.01
The bloom is off the rose	7	0.01
Full of years	7	0.01

Table 5 Thematic index of Oxford Dictionary of Idioms along with the frequencies per million *(Continued)*

Idiom	Frequency	Frequency per million
Pass your sell-by date	5	0.01
Put years on someone	4	0.01
Threescore years and ten	3	0.01
There's many a good tune played on an old fiddle	0	0
The vale of years	0	0
Stricken in years	0	0
Have had a good innings	0	0
Ambition		
Set your sights on	461	0.87
Fly high	392	0.74
Think big	325	0.62
Fire in the (or your) belly	113	0.21
Reach for the stars	64	0.12
Room at the top	48	0.09
Bite off more than you can chew	46	0.09
Punch your ticket	44	0.08
Raise (or lower) your sights	41	0.08
Set your heart (or hopes) on	40	0.08
Punch above your weight	22	0.04
Run before you can walk	0	0

Acknowledgements
Not applicable.

Authors' contributions
ER conducted the study under the supervision of AA. All the steps of the research were done collaboratively. All authors read and approved the final manuscript.

Authors' information
Elaheh Rafatbakhsh is a PhD candidate of TEFL in the Department of Foreign Languages and Linguistics at Shiraz University. Her main research interests include Computer Assisted Language Learning, Language Assessment and Corpus Linguistics. Alireza Ahmadi is an associate professor of Teaching English as a Foreign Language in the Department of Foreign Languages and Linguistics at Shiraz University, Shiraz, Iran. He teaches second language assessment and language learning courses. His main area of research is second language assessment.

Funding
The authors received no financial support for the research, authorship, or publication of this article.

Availability of data and materials
The datasets analyzed during the current study are available in https://corpus.byu.edu/coca/
The datasets generated in the current study are available from the corresponding author on reasonable request.

Competing interests
The authors declare that they have no competing interests.

Received: 10 December 2018 Accepted: 20 August 2019
Published online: 17 October 2019

References

Alexander, R. J. (1987). Problems in understanding and teaching idiomaticity in English. *Anglistik und englischunterricht, 32*, 105–122.
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* (p. 87). London: Continuum Press.

Anthony, L. (2012). *AntConc (version 3.3.5) [Computer software]*. Tokyo: Waseda University Available from http://www.antlab.sci.waseda.ac.jp/.
Baddorf, D. S., & Evens, M. W. (1998). Finding phrases rather than discovering collocations: Searching corpora for dictionary phrases. In *Proc. of the 9th Midwest Artificial Intelligence and Cognitive Science Conference (MAICS-98)* (pp. 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.
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. (2000). Metaphor awareness and vocabulary retention. *Applied Linguistics, 21*(4), 553–571.
Boers, F., Eyckmans, J., Kappel, J., Stengers, H., & Demecheleer, M. (2006). Formulaic expressions and perceived oral proficiency: Putting a lexical approach to the test. *Language Teaching Research, 10*(3), 245–261.
Boers, F., & Stengers, H. (2008). Adding sound to the picture: Motivating the lexical composition of metaphorical idioms in English, Dutch and Spanish. In M. S. Zanotto, L. Camron, & M. C. Cavalcanti (Eds.), *Confronting metaphor in use* (pp. 63–78). Amsterdam: John Benjamins B.V.
Boers, F. (2013). Cognitive linguistic approaches to teaching vocabulary: *Assessment and integration. Language Teaching, 46*(02), 208–224.
Cacciari, C. (1993). The place of idioms in a literal and metaphorical world. In C. Cacciari & P. Tabossi (Eds.), *Idioms, processing, structure and interpretation* (pp. 76–87). Hillsdale: Lawrence Erlbaum Associates.
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.
Cameron, L., & Low, G. (1999). Metaphor. *Language Teaching, 32*, 77–96.
Chafe, W. L. (1968). Idiomaticity as an anomaly in the Chomskyan paradigm. *Foundations of Language, 4*, 109–127.
Cooper, T. C. (1999). Processing of idioms by L2 learners of English. *TESOL Quarterly, 33*(2), 233–262.
Cowie, A. P., Mackin, R., & McCaig, I. R. (1983). *Oxford dictionary of current idiomatic English: Phrase, clause and sentence idioms (Vol. 2)*. Oxford: Oxford University Press.
Crystal, D. (1997). *English as a global language*. Cambridge: Cambridge University Press.
Daskalovska, N. (2011). The impact of reading on three aspects of word knowledge: Spelling, meaning and collocation. *Procedia - Social and Behavioral Sciences, 15*, 2334–2341.
Davies, M. (2015). The corpus of contemporary American English: 520 million words, 1990-present. Available from https://www.english-corpora.org/coca/
Ellis, N. C. (1994). Vocabulary acquisition: The implicit and outs of explicit cognitive mediation. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages*. London/San Diego: Academic.
Fernando, C. (1996). *Idioms and idiomaticity*. Oxford: Oxford University Press.
Fraser, B. (1970). Idioms within a transformational grammar. *Foundations of Language, 6*(1), 22-42.
Glucksberg, S., & McGlone, M. S. (2001). *Understanding figurative language: From metaphor to idioms* (Oxford psychology series, no. 36). New York: Oxford University Press.
Gramley, S., & Pátzold, M. (2003). *A survey of modern English*. New York & London: Routledge.
Grant, L. E. (2003). *A corpus-based investigation of idiomatic multiword units* (Doctoral dissertation). Retrieved from http://researcharchive.vuw.ac.nz/xmlui/bitstream/handle/10063/327/thesis.pdf?sequence=2
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? *Lang & Ling Compass, 3*, 1–17.
Hardie, A. (2013). *CQPweb* [Computer Software]. Available from http://cwb.sourceforge.net/cqpweb.php.
Johansson, S. (2009). Some thoughts on corpora and second-language acquisition. In K. Ajimer (Ed.), *Corpora and language teaching* (pp. 33–44). Amsterdam: John Benjamins.
Kilgariff, A. (2013). *SketchEngine* [Computer Software]. Available from http://www.sketchengine.co.uk/.
Kovecses, Z., & Szabco, P. (1996). Idioms: A view from cognitive semantics. *Applied Linguistics, 17*(3), 326–355.
Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning, 54*, 399–436.
Lewis, M. (1993). *The lexical approach*. Hove: 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: 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.
Mauranen, A. (2004). Spoken corpus for an ordinary learner. In J. M. Sinclair (Ed.), *How to use corpora in language teaching* (pp. 89–105). Amsterdam/Philadelphia: John Benjamins Publishing Company.
Moon, R. (1998a). Frequencies and forms of phrasal lexemes in English. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis, and applications* (pp. 79–100). Oxford: Clarendon Press.
Moon, R. (1998b). *Fixed expressions and idioms in English: A corpus-based approach*. New York: Oxford University Press.
Muller-Hartmann, A., & Schocker-von Ditfurth, M. (2004). *Introduction to English language teaching*. Stuttgart: Klett.
Nation, I. S. P., & Meara, P. (2002). Vocabulary. In N. Schmitt (Ed.), *An introduction to applied linguistics* (pp. 35–54). New York: Oxford University Press.
Nattinger, J. (1988). Some current trends in vocabulary teaching. *Vocabulary and language teaching, 1*, 62–82.
O'Dell, F., & McCarthy, M. (2010). *English idioms in use (advanced)*. Cambridge: Cambridge University Press.
Ovando, C. J., & Collier, V. P. (1985). *Bilingual and ESL classrooms*. New York: McGraw-Hill Book Company.
Panou, D. (2014). *Idioms translation in financial press: A corpus-based study*. Newcastle upon Tyne: Cambridge Scholars Publishing.
Schmitt, N. (2004). Formulaic sequences: *Acquisition, processing, and use (Vol. 9)*. Amsterdam, Philadelphia: John Benjamins.
Simpson, R., & Mendis, D. (2003). A corpus-based study of idioms in academic speech. *TESOL Quarterly, 37*(3), 419-441.

Scott, M. (2012). *WordSmith Tools (Version 5.0)* [computer software]. Available from http://www.lexically.net/software/index.htm.
Szczepaniak, R. (2006). *The role of dictionary use in the comprehension of idiom variants*. Tübingen: Niemeyer.
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.
Tran, H. Q. (2013). Figurative idiomatic competence: An analysis of EFL learners in Vietnam. *Language Education in Asia, 4*(1), 23–38.
Tsui, A. B. (2004). What teachers have always wanted to know–and how corpora can help. *How to use corpora in language teaching, 12*, 39–61.
Vaclavikova, E. (2010). *Idioms of colour – A Corpus-based study* (Master's thesis, Masaryk University). Retrieved from https://is.muni.cz/th/z0q9v/diplomka_IS.pdf.
Vrbinc, A., & Vrbinc, M. (2011). Creative use of idioms in satirical magazines. *Jezikoslovlje, 12*(1), 75–91.
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.
Wood, M. M. (1981). *A definition of idiom*. Bloomington: Indiana University Linguistics Club.
Wray, A. (2000). Formulaic sequences in second language teaching: Principle and practice. *Applied Linguistics, 21*(4), 463–489. https://doi.org/10.1093/applin/21.4.463.

Publisher's Note