

The Most Frequent Opaque Idioms in English News

Wenhua Hsu

I-Shou University, Kaohsiung, 84001, Taiwan

Email: whh@isu.edu.tw

Abstract

This research aimed to establish a pedagogically useful list of the most frequent opaque idioms in English news. It began by compiling an idiom search list from six prestigious idiom dictionaries. Through a set of criteria, 4,864 semantically non-compositional idioms were culled as search entries to interrogate the News on the Web (NOW) Corpus—the largest news corpus to date. A total of 525 most frequent opaque idioms were ultimately selected. To verify if they merit pedagogical concern, the 525 idioms were tested on the Voice of America (VOA) News Corpus. Results show that they accounted for 0.59% and 0.61% of running words of the NOW and VOA Corpora respectively. Despite a small percentage, knowledge of opaque idioms may contribute to filling the rift of lexical coverage that individual words fail to account for in news articles. For English learners, this opaque idiom list provides a window to the vast number of idioms used in daily news and can serve as a reference in setting lexical goals at the initial phase of idiom learning.

Keywords: non-compositional; NOW Corpus; idioms; lexical coverage

Introduction

Within English language learning, one challenge for learners is the huge number of idioms. A mastery of idioms is often regarded as native speaker fluency (Fernando, 1996; Schmitt, 2000; Simpson & Mendis, 2003). In journalistic register, idioms may be more extensively used than any other discourse registers. For instance, in news headlines, idioms may be manipulated to achieve certain effects such as irony or humor to intrigue readers. In accounts of events, journalists often use idioms as a shorthand way of presenting their points crisply when idioms can provide images of what is being said, e.g., *call the shots*, *jump the gun*, *rock the boat*, *out of pocket*, *put on the back burner*, *take a back seat*, *the upper hand*, *win hands down*, to name but a few.

Given the vast inventory of idioms in a native speaker's lexical repertoire, deciding which idioms should be taught during limited class sessions is a challenge for most English teachers. Although contemporary English idiom dictionaries are of great help for idiom learning, they may contain a substantial number of seldom-used or archaic idioms (Liu, 2003). English language teaching (ELT) publishers often claim that their textbooks contain essential idioms, but according to Chen and Wang's (2016) survey on ELT materials, the so-called essential idioms are often selected based on the author's intuitive judgement rather than from empirical evidence. As a result, some textbook-selected idioms are rarely used in real language situations.

The issue of making principled decisions about which idioms are worth focusing on can be addressed through a corpus-based approach. Biber, Johansson, Leech, Conrad and Finegan (1999) advocated that a large collection of naturally-occurring language data can provide a rigorous way for identification through frequency. Not all English idioms are equally important. In terms of a good learning return, targeting a restricted number of idioms with relatively high frequency of occurrence may be more practical in this regard.

Using a corpus-based approach, Simpson and Mendis (2003) uncovered 238 idiom types from the 1.7-million-plus-word Michigan Corpus of Academic Spoken English (MICASE). Liu

(2003) compiled a list of idioms from four English idiom dictionaries and three phrasal verb dictionaries. Then he engaged in a laborious concordance search in three contemporary spoken American English corpora plus one self-compiled media corpus and identified 302 most frequent spoken American English idioms. Using empirical data sets, both studies have helped English instructors to decide which spoken idioms should be taught first.

Different from Simpson and Mendis (2003) as well as Liu (2003), this research targeted news texts as a corpus source for two reasons. First, news articles provide coverage of current affairs. Each news report is real-life and may reveal what is trending today, so the recurrent idioms selected from within would reflect how often they are likely to be encountered in daily life and evidence that they are not outdated idioms. Second, news articles are first and foremost indispensable learning material for English for Journalism courses. A list of high-frequency idioms directly derived from news articles may immediately meet the lexical needs of the user. Also, different from previous studies, this research was more concerned with opaque idioms. Here opaque idioms refer to semantically non-compositional idioms, of which the individual words do not help each other to reveal the meaning as a whole.

When multiword expressions are composed of known words and their meanings as a whole turn to be unfamiliar to learners, lexical coverage (the percentage of known words to the total) associated with comprehension may thus be overestimated in this regard. A case in point is opaque idioms, particularly consisting of high-frequency general words (e.g. *a can of worms*, *someone's cup of tea*, *white elephant*, *an arm and a leg*, *under one's belt*). As shown, these idioms usually do not mean what they literally state and cannot be interpreted word for word. This study narrowed the research scope to opaque idioms, because they may pose comprehension hurdles if they are not known. Consequently, the purpose of this research was to identify the most frequent opaque idioms in English news, which are worthy of pedagogical attention. This research sought to answer the following three questions.

1. What are the most frequent opaque idioms in English news?

2. What discourse functions do opaque idioms perform in English news?
3. What is the text coverage of the most frequent opaque idioms in English news?

Literature Review

Idiom definitions and types

Moon (1998) defined idioms as those “fixed and semantically opaque or metaphorical” expressions (p.4), whereas Fernando (1996) outlined them as “conventionalized multiword expressions often, but not always non-literal” (p. 1). Both viewpoints were integrated by Cooper (1998), who compared idioms to metaphors and concluded that an idiom can have a literal meaning, but its alternate, figurative meaning must be understood metaphorically.

Not all idioms are equally opaque in meaning. Based on literality, Fernando (1996) divided idioms into three categories: pure (non-literal), semi-literal and literal idioms. In a similar vein, Grant and Nation (2006) pointed out that there are three types of idioms: core, figurative and literal. In contrast, Glucksberg (2001) identified four types of idioms according to semantic compositionality: non-compositional, compositional opaque, compositional transparent and quasi-metaphorical. Each type shows the relationship between an idiom’s constituents and its meaning.

As far as literality is concerned, the division between literal and non-literal idioms is often blurred (McCarthy, 1998). Some idioms have both a literal and a non-literal meaning, subject to context. For instance, *bread and butter* may literally mean food items or have a metaphorical meaning for a living; *red tape* may refer to red ribbon used to tie things or allude to excessive official regulation. Other examples include *silver bullet*, *olive branch*, *a house of cards* and so on. Moon (1998) found that the literal meanings of such idioms are always rarer than their idiomatic interpretation.

Different from Fernando (1996), Grant and Bauer (2004) used three criteria, non-compositionality, figurativeness and ONCE to divide a large collection of idioms into three groups: core idioms

(non-compositional but non-figurative), figuratives (non-compositional and figurative) and ONCEs (one non-compositional element, possibly figurative). Non-compositionality refers to semantic opacity, which denotes that the entire meaning of an idiom cannot be predicted by analyzing the meaning of its component words. Figurativeness involves the recognition of untruth. The meaning of a figurative idiom needs to be interpreted non-literally. ONCEs refer to the idioms with only one word that is either non-literal or non-compositional.

Later, Grant and Nation (2006) applied these three criteria to all of the entries in ten English idiom dictionaries and identified a total of 103 core idioms. Using the 103 core idioms, Grant (2005) conducted a concordance search in the British National Corpus. The search revealed that none of them occurred frequently enough to merit inclusion in the first 5,000 word families. This outcome echoed the earlier studies on pure idioms. Searching in the Oxford Hector Pilot Corpus (OHPC) and Bank of English (BoE) Corpus, Moon (1998) discovered that pure idioms are very rare across the board. In an analysis of the Longman Spoken and Written English Corpus exceeding 40 million words, Biber et al. (1999) also found pure idioms to be scanty—generally less than one token per million words.

Of several criteria that have long been used to define an idiom, the most commonly accepted criterion is non-compositionality. Another two defining criteria for idioms are fixedness and institutionalization. Idioms constitute set expressions and their component words cannot be easily substituted for. Institutionalization refers to the degree of acceptance among the wider discourse community. A fixed expression is considered to be an idiom when it is widely recognized and familiar (Fernando & Flavell, 1981). Despite the well-established criteria, lack of precision in defining an idiom has resulted in its proliferation encompassing a variety of idiom types.

Table 1. Past studies on idiom types

Makkai, 1972	Alexander, 1987	McCarthy, 1998
Phrasal verbs	Phrasal verbs	Phrasal verbs
Turns of phrase	Turns of phrase	Turns of phrase
Irreversible binomials	Irreversible binomials	Irreversible binomials
Phrasal compounds	Phrasal compounds	Prepositional
Pseudo-idioms	Proverbial idioms	expressions
Proverbs	Metaphorical/allusive	Frozen similes
Familiar quotations	idioms	Possessive's phrases
Idioms associated with	Idiomatic similes	Opaque compounds
a national game		Idiomatic speech
Institutionalized		routines
expressions		Restricted collocations
		Cultural allusions

As can be seen in Table 1, in a broader sense, some researchers consider phrasal verbs, proverbs, similes, binomials and prepositional expressions as idioms, while some researchers (Cooper, 1998; Glucksberg, 2001) even view a restricted collocation or a verb plus a noun that is used figuratively as an idiom (e.g. *cheesecake* photos; *weigh* a decision). Therefore, the boundary between idioms and non-idioms is often ambiguous.

Past studies reveal that there has hitherto been little consensus among researchers on what constitutes an idiom (Liu, 2003). As far as this research is concerned, it may be futile to identify idioms from scratch, and therefore it resorted to prestigious idiom dictionaries and used the idioms from within as search entries.

Discourse functions of idioms

Among the discourse functions identified in the literature, relevant to the present research are Biber, Conrad, and Cortes (2004) as well as Simpson and Mendis (2003). Some of the discourse functions they identified from lexical bundles and spoken idioms are also well suited to opaque idioms, since they are all multiword expressions and exhibit one facet of formulaic language.

Biber, Conrad, and Cortes (2004) designed a categorization scheme for lexical bundles commonly used in university spoken

and written registers. In their taxonomy, there were four core categories: referential, discourse-organizing, stance and interactional bundles. Referential bundles make direct reference to textual context and the signals they send include time, location, procedure, quantity and the description of attributes. Discourse-organizing bundles are concerned with the organization of a text, indicating inferential, contrastive or causative relations or signaling transition. Stance bundles convey the writer's attitude or assessments of certainty, which provide a frame for the interpretation of subsequent propositions, while interactional bundles refer to the ways the writer intervenes to actively address readers.

Concerning the discourse functions of spoken idioms, Simpson and Mendis (2003) enumerated six primary functions based on 238 idiom types they identified from the MICASE. They illustrated certain idioms in MICASE that are used for evaluation, description, paraphrase, emphasis and collaboration as well as used in metalanguage. Simpson and Mendis (2003) discovered that evaluative uses of idioms are often descriptive, but descriptive uses do not always entail evaluation. They also found that paraphrasing uses of idioms often have the effect of reducing the distance between the speaker and listener through the juxtaposition of a professional jargon with a more colloquial idiom. In academic discourse, speakers use idioms to reinforce an explanation. Moreover, idioms that perform the collaborative function often appear when participants in a conversation express their shared views. As to the use of idioms in metalanguage, they function as signposting and organizational devices to create coherence and intelligibility in a discourse.

The above categorization was helpful to this research in identifying the functions of idioms used in news articles. However, a single idiom may perform more than one discourse function and a clear-cut distinction of them may be fruitless. Accordingly, this research targeted their main functions in news articles.

Method

The NOW Corpus

The NOW (News on the Web) Corpus (<https://www.english-corpora.org/now/>) is the largest, publicly-available, well-balanced English news corpus to date. At the time of undertaking this research, it has already had 9.69 billion words of data retrieved from web-based newspapers from 2010 to the present. Automated scripts run every day to add texts to the corpus. Each day these scripts get the URLs from Google News, so the corpus is continually growing by 140 to 160 million words each month. Automatic data expansion has important implications for idiom use, since the very low frequencies in the NOW Corpus may indicate that the idioms have actually gone out of fashion and thus may be of little pedagogical value (e.g. *a fly on the wheel*, not even appearing once).

Furthermore, 6 million words of Voice of American (VOA) news were downloaded from <https://www.voanews.com/> for validity check (for VOA news retrieval, see Hsu, 2019).

Compiling a search list

Although some reputable idiom dictionaries are claimed to be comprehensive in containing essential idioms, the selection varies from dictionary to dictionary. For example, Longman Dictionary of English Idioms include *after all*, *above all*, *take into account*, *by virtue of* in it while Oxford Dictionary of Idioms does not. To settle such discrepancy, an idiom was included in the search list if it was listed in at least three of these six dictionaries: Farlex Idioms and Slang Dictionary (2017), Collins COBUILD Idioms Dictionary (2012), Oxford Idioms Dictionary for Learners of English (2007), McGraw-Hill's Dictionary of American Idioms and Phrasal Verbs (2006), The American Heritage Dictionary of Idioms (2013) and Longman Idioms Dictionary (1998). Moreover, two online dictionaries, Cambridge International Dictionary of Idioms-iTools (2019) and The Free Dictionary-idioms (2019) were consulted. This step ensured the selection of idioms with wide acceptance and recognition. Through the alphabetically sorting function in Excel, repeated idioms among the six dictionaries were removed. Totally

11,983 different idioms were identified with some needing compiling.

As aforementioned, this research was more concerned with opaque idioms that are composed of the most frequent words and may be concealed in known words. If no distinction is made between very common words and the idioms containing them, the latter may go overlooked or misinterpreted (e.g. *a far cry from, pull one's leg, be that as it may, the likes of, out of the blue*). Accordingly, the idiom search list was decided to stop at high-frequency words. Here high-frequency vocabulary refers to the first 3,000 word families along the word-frequency scale of the BNC (British National Corpus) and COCA (Corpus of Contemporary American English) as well as the General Service List of English (GSL) (West, 1953). The decision was admittedly arbitrary but not utterly unrealistic. English learners who are encouraged to read English news should already have receptive knowledge of the most frequent words. From a pedagogical viewpoint, we hope our students are able to learn idioms with a minimum of difficulty at the onset of idiom learning. They may be distracted by unknown words if there is one in an idiom. Selecting idioms composed of familiar words (e.g. *across the board, touch base, worth one's salt, with/without a grain of salt*) may be a good start to approaching idioms when students are willing to give English news reading a shot.

To examine the vocabulary levels of the 11,983 idioms, the AntWordProfiler (Anthony, 2014) was used to run the BNC/COCA word lists (Nation, 2017) over the idioms. Based on frequency, range and dispersion, the ranked BNC/COCA twenty-five 1,000-word-family lists provide an estimate of the vocabulary level of a word or a text. At this screening stage, a cut-off for exclusion of idioms with component words beyond the BNC/COCA 3000 and the GSL (e.g. *tilt at windmills, a whited sepulcher, in the doldrums*) resulted in a nearly two-fifths reduction from 11,983 to 7,190 idioms.

In contrast with Grant and Bauer's (2004) distinction of core idioms, figurative idioms and ONCEs, this research considered the three types as opaque idioms, since they all contain at least one word that is semantically non-compositional and may lead to

deceptive comprehension (learners think they know but actually they do not) (Martinez & Murphy, 2011). Finally, to judge whether the remaining items were opaque idioms, this research formulated two statements to guide the decision for inclusion in the search list.

S1: The meaning of the idiom as a whole does not remain or marginally remains when each constituent part is decoded with its key meaning.

S2: The idiom contains one of the meanings that its component word cannot account for.

The researcher and two colleagues made an independent judgement. When either of the two statements was true for an idiom, it was regarded as an opaque idiom. When there was no agreement among the raters (Cohen's kappa $k = 0.86$), the idiom was reserved for further analysis until a consensus was reached. After manual vetting, 4,864 opaque idioms were '*waiting in the wings*' for subsequent corpus search.

Determining a cut-off point

The NOW Corpus has its own search interface, where the 4,864 opaque idioms were used as search queries. The next decision was what frequency level was to be used as a cut-off. For a single word to enter the BNC first 5,000 word families, the word and its family members altogether need to occur at least 7.87 times per million words (Nation, 2005). However, frequencies decrease as single words are expanded to multiword units and their frequencies drop drastically as multiword combinations are extended to five words or beyond (Hyland, 2008). Predictably, idioms containing more than one word are far less frequent than their component words (*touch wood* appearing 1,505 times in the NOW Corpus versus *touch* 572,863 times and *wood* 296,210 times) and long idioms become even rarer (*there are more ways than one to skin a cat* appearing only 6 times; *six of one and half a dozen of the other* 41 times).

Moon (1998) divided the lowest band of medium-frequency idioms at a frequency of twice per million words. Biber et al.

(1999) noted that very few pure idioms appear greater than once per million words. In this research, the pilot study on the 103 core idioms (Grant & Nation, 2006) showed that *eat crow* and *give someone the bird* occurred only 289 and 125 times in the 9.69-billion-word NOW Corpus and *wear the green willow* did not even appear a single time. Nevertheless, the core idiom *by and large* occurred 21,048 times (2.17 times per million words). The average frequency of the 103 core idioms in the Now Corpus was 2,502 times (roughly 0.26 times per million words). Based on the average frequency, the cut-off point was arbitrarily set at 2,500 times. In contrast with Moon's (1998) dividing point at twice per million words for mid-frequency idioms (including literal and compositional idioms), the lower cut-off point was used to avoid the omission of pure/core idioms (non-compositional and non-figurative).

Challenging a corpus search

Fixedness is a key property of idioms. However, as many as 40% of idioms have some lexical variations or may undergo institutionalized transformations, including the permutation, deletion or addition of component words (Moon, 1998). The variations of an idiom involve verb inflection, passivization, pronoun or possessive changes, apostrophe, participle variation, truncation, hyphenation and plural/singular forms (Grant, 2005). Thus, entering various possible forms of an idiom in the query box on the NOW interface became a challenge. The researcher endeavored to make a laborious search by recording an exhaustive list of various possible forms at her best.

Because the canonical form of an idiom and its variants share the same core meaning, their occurrences should be counted together so that the combined frequency would indicate whether the idiom was salient enough to be considered for inclusion. The separate counting of each form may cause the idiom to be filtered out, when the individual frequency is insignificant but the combined frequency is significant [e.g. 'to a tee/T' meaning 'exactly or to perfection' with a sum of 3,279 times = *to a tee* (1,837 times < the cutoff 2,500 times) + *to a T* (1,442 times < 2,500)]. Another

example is *throw/have/pitch a fit*, meaning very angry, often shouting a lot [a sum of 2,042 times = *throw* (*throws, threw, thrown, throwing*) *a fit* (771 times) + *have* (*has, had, having*) *a fit* (1,100) + *pitch* (*pitches, pitched, pitching*) *a fit* (171)]. Because 2,042 times were still below the cutoff 2,500 times, the idiom *throw/have/pitch a fit* was thereby excluded. For more details concerning how an idiom was searched for in the NOW to obtain the total frequency, see Table 2.

Table 2. Search examples in the NOW Corpus

Canonical form	Search syntax	Resulting types of form	Total freq.
follow suit	[follow] suit	follow, follows, followed, following suit	46,304
cut one's teeth on/in/at	[cut] _app* teeth on in at [or use a drop-down parts-of-speech (POS) menu for poss.ALL]	I cut my teeth on; she cut her teeth at...	6,932
tongue in cheek	tongue-in-cheek tongue in cheek	tongue-in-cheek tongue in cheek	12,634
take/call someone to task	[take] * to task [take] ** to task [take] *** to task [call] likewise.	take, takes, took, taking call, calls, called, calling him/them/John...to task	3,317
a bird in the hand	a bird in the *	a bird in the hand	208
	a bird in the *	a bird in the bush, sky, air or wild...	152
at sea	[be] at sea [be] * at sea [be] ** at sea [be] *** at sea [feel] likewise.	am, are, is, being, was, were, been feel, feels, felt, feeling completely/totally/rather/all/ somewhat/a little bit/a bit/ a little/ kind of/sort of... at sea	2,504
a piece of cake	a piece of cake	a piece of cake (cake for real) 6%* 2,503=150 a piece of cake (easy task) (1-6%)* 2,503=2,353	2,503

Canonical form	Search syntax	Resulting types of form	Total freq.
no mean + noun	no mean _nn* (or use a drop-down POS menu for noun.ALL)	~is no mean trick. ~ was no mean achievement. ~ is no mean player.	6,971

Note: The wildcard character * represents any single token; | = alternant; [] = word lemmas.

Table 2 demonstrates the search syntax for various forms of an idiom and its total frequency (on the right column). The use of search syntax can reduce the number of entries in the search bar while bringing about the same result. For example, the square brackets were used for the lemma search of the verb *follow* in the idiom *follow suit*, meaning ‘to do the same thing’. Each form (*follow*, *follows*, *followed*, *following*) was searched for separately and all the frequencies were automatically added up.

The idiom *cut one’s teeth* is often followed by a preposition *on*, *in* or *at*, but the three forms were considered as one idiom due to the same core meaning. The symbol | signifies the alternant search for the three prepositions. The NOW engine can calculate their respective frequencies and the total frequency. Furthermore, a possessive adjective often appears between *cut* and *teeth*. The query string *_app** for a possessive adjective can be typed in the search bar or a drop-down parts-of-speech (POS) menu for *poss.ALL* can be used.

Moreover, there were variations in hyphenation. The idiom *tongue in cheek* has a hyphenated form and a non-hyphenated form. The occurrences of both forms were summed up to determine whether the frequency of the idiom was high enough for consideration.

Another example is ‘*take someone to task*’, meaning ‘reprimand someone severely for a fault’. There are two variable slots for this idiom. One is the verb *take* and the other is the object *someone*. This idiom was entered in the search bar as [take] * to task, with [take] instructing the search engine to look for the lemmas of the verb *take* (*takes*, *took*, *taken*, *taking* and *take*) and with *someone* being replaced by the asterisk wildcard * denoting any token between *take* and *to task*. As a result, the use of the

asterisk in the search bar is highly likely to give rise to a variety of forms, as [take] him/John/proper nouns to task have shown.

By the same token, the truncated idiom *a bird in the hand* is presented in the search syntax as a bird in the *. The variants *a bird in the hand* and *a bird in the bush* as well as *a bird in the sky*, *air* or *wild* were thus generated. The meaning of a *bird in the bush*, *sky*, *air* or *wild* is different from that of *a bird in the hand* or the full form *a bird in the hand is worth two in the bush*, so its frequency was not added to that of the latter two.

The idiom *at sea* is found to be used in both senses, meaning puzzled or sailing on the sea. Based on 100 random concordance lines, the ratio of non-literal use to literal use was one to one (see *a piece of cake* below for further explanation). To distinguish its non-literal sense from literal one, the copulas *be* and *feel* were added. The idiom also appears with adverbs often, such as *all*, *totally*, *completely*. In view of these instances, the search syntax was typed in as [be] * at sea and [feel] * at sea.

Like *at sea*, the idiom *a piece of cake* can carry either a literal or a non-literal meaning, subject to contexts. The former refers to edible cake while the latter means something easily achieved. Therefore, the concordances of *a piece of cake* were reviewed to eliminate the instances of the literal meaning. In a gigantic corpus, line-by-line scrutiny was not a viable option. In this research, only 100 concordance lines were checked manually. The proportion of the idiomatic/non-literal use was calculated and then converted to percentage (frequency of non-literal use divided by the total frequency of both literal and non-literal use). To validate the percentage, another 100 concordance lines were reviewed for reliability check. In rare cases, additional random samples were needed until a reliable percentage figure could be approximated. Table 2 shows that *a piece of cake* occurred 2,503 times in the NOW and the percentage of its literal use was 6%. Based on this percentage, it was extrapolated that its non-literal meaning occurred approximately 2,353 times in the corpus [(1-6%)* 2,503=2,353].

When searching the idiom *no mean* (denoting a very good), some parts of speech such as nouns need to be specified. In this situation, a drop-down parts-of-speech (POS) menu on the NOW

interface is easier to use than typing a query. The search syntax will automatically show up in the search bar when a certain part of speech is chosen from the drop-down menu. Then all the respective frequencies were added up.

Results and Discussion

The most frequent opaque idioms in English news

A total of 525 opaque idioms within the first 3,000 word-family level and the GSL were ultimately chosen and formed the most frequent opaque idiom list (see Appendix and see next section for their functions). Table 3 presents the composition of the idioms concerning vocabulary levels. The opaque idiom list consists of 1,994 running words and involves 594 word families.

Table 3. Lexical coverage at each 1000 of the BNC/COCA 3000 for the 525 opaque idioms

BNC/COCA base word lists	Tokens	% coverage in tokens	Cumulative % coverage	Number of word families
1 st 1000	1,482	74.32%	74.32%	371
2 nd 1000	284	14.24%	88.56%	134
3 rd 1000	132	6.62%	95.18%	49
GSL outside the BNC/COCA 3000	94	4.71%	99.89%	38
Proper nouns (Note below)	2	0.11%	100%	2
Total	1,994	100%		594

Note: *Achilles' heel* and *Pandora's box* were included because *Pandora* and *Achilles* are proper nouns. It was assumed that a place name or a personal name can be recognized without much effort. 'GSL outside the BNC/COCA 3000' refers to the words in the GSL that do not overlap with the BNC/COCA 3000.

As is evident in Table 3, a majority of the 525 idioms are comprised of the BNC/COCA first 1,000 word families. The first 1,000 word families account for 74.32% of the total words in the idiom list and the second 1,000 make up 14.24%, with the percentage of the third 1,000 being 6.62% and that of the GSL outside the BNC/COCA 3000 being 4.71%.

Setting the screening criterion at the BNC/COCA 3000 and the GSL, the opaque idiom list contains many topic-neutral idioms composed of very general words (e.g. *on the ball*, *call it a day*, *have one's work cut out*). Along with their high-frequency component words, the 525 opaque idioms occur across a wide range of news topics. Examples include *walks of life*, *face the music*, *as and when*, *state-of-the-art*, to name but a few.

Major discourse functions of opaque idioms in English news

There are four basic functions that a newspaper (print or online) can offer to its readership. Hard news that reports specific events aims to inform readers. Editorials, commentaries and reviews provide further interpretations of what was happening to fill the gaps left by hard news. Features and columns serve readers to keep abreast of knowledge of modern life in all aspects. In addition to comic columns that give enjoyment to readers, *off-the-cuff* idioms along with witty remarks in texts amuse readers as well.

McCarthy (1998) pointed out that idioms are highly interactive expressions and cannot be used without motivation. Considering the functions of a newspaper (to inform, to interpret, to serve and to entertain) and referring to Biber, Conrad and Cortes (2004) (referential, discourse-organizing, stance and interactional bundles) as well as Simpson and Mendis's (2003) taxonomy of spoken idioms (evaluation, description, paraphrase, emphasis, collaboration and metalanguage), the researcher generalized idiom use in news texts into five types: (1) referential uses (including informing and interpreting purposes such as quotes from eyewitness accounts), (2) evaluative uses, which are what McCarthy (1998) referred to as the observation-plus-comment function, such as debates of stances, moral judgements, (3) emphasizing uses (e.g. presenting a contrast of opinions, highlighting content, reinforcing an explanation), (4) discourse-organizing uses for coherence and intelligibility, and (5) word play. To achieve amusing or ironic effects, idioms are manipulated in a punny way. Such word play involves a conflict between its literal meaning and non-literal meaning. The following news excerpts show some cases of the above.

Referential uses

A statement from Australia's Defense Ministry said it would monitor the "air situation in Syria" and make a decision on resuming airstrikes there in good time.

Reuters

U.S. officials say their focus is on defeating IS in Syria and Iraq, and that they expect their allies to all be on the same page.

Military Times

Evaluative uses

It is high summer now and some Italian families who've maintained their seaside homes in Castel Volturno against the grain of history present an incongruous picture trooping toward Death Triangle's beach in their swimwear with kids clutching buckets and spades.

VOA News

"Proposing to extend the 14-day limit might be opening a can of worms, but would it lead to Pandora's box, or a treasure chest of valuable information?"

The Telegraph

Referential and evaluative uses

"The line of diplomacy or inter-Korean relations should be interpreted in terms of North Korean interests as kind of a tactic to buy time," said Thae Yong Ho, Pyongyang's former deputy ambassador to Britain.

Time

"It's normal and reasonable to react to short-term worries, but in the long run, any damage from the election is likely to be limited," said Brad McMillan, chief investment officer at Commonwealth Financial Network.

Forbes

Emphasizing uses

"These results are not coming from out of the blue, they are marked by facts, and you cannot claim that results are fake with respect to presidential and you welcome the areas where your governors and your members of parliament have won convincingly."

New Delhi Times

Discourse-organizing uses

The factual reality is that Umzimkhulu is both rural and very vast, so service delivery backlogs, especially for electricity, are quite problematic. On that note, community engagements are justifiably robust in most cases.

News24

Word play for fun or sarcasm

Fat cat bosses under fire—According to the Mail on Sunday, the prime minister has launched an outspoken war on fat cat bosses - executives whose salaries outstrip the performance of their companies.

BBC

Note: Next to this news report is the comic strip of a very fat cat being shot.

As can be seen above, idioms are multifaceted and it may not be easy to fold them into a compact categorization. Table 4 provides an overall distribution of the idioms across the five discourse functions. Among the 525 opaque idioms, there are about 183 referential idioms, 178 idioms used for evaluative purposes, 81 idioms for emphasis, 52 idioms for word play and 31 discourse-organizing idioms. Although the idioms were classified based on their most common use, it may be challenged on the precision of categorization as a result of multiple functions subject to context. This preliminary typology was used to explore a general pattern concerning the usage of idioms in news articles. Results show that a high proportion of opaque idioms in the NOW are referential and evaluative idioms, with both accounting for 69% of the total idioms $[(183+178)/525]$. The idioms for emphasis are the thirdly dominant (15%) followed by those for entertaining purposes (10%). By contrast, discourse-organizing idioms are far less common (6%) in journalistic discourse.

Table 4. Distribution of the opaque idioms across five functions

Pragmatic functions	Number of idioms	Instances
Referential uses	183	catch-22; the order of the day; an ace up one' sleeve; keep ~ at bay; sweep ~ under the carpet; put paid to; benefit of the doubt; the likes of
Evaluative uses	178	elephant in the room; play the game; icing on the cake; under the skin; make or break; over the top; hold water, outside the box
Emphasizing uses	81	with flying colors; eleventh hour; not by a long shot; few and far between; make no mistake; to death; in broad daylight
Word play	52	pork barrel; a house of cards; spill the beans; throw in the towel; pull the plug; under the hood; on one's plate; fat cat
Discourse-organizing uses	31	for good measure, on that score; in the light of; in a nutshell; to all intents and purposes; by and large; be that as it may

This may be ascribed to the reason that the dissemination of information is crucial to the profession of journalism. Journalists strive to make news reports engaging and intelligible to the majority of readers and attempt to answer the 5W1H questions about any particular event. Apart from informing purposes, elements of appraisal in news articles can also be observed in the very selection of idioms. This is because idioms sometimes bring a clear mental picture to the mind of the reader and news writers use them to express something more vividly in controversial or high-profile issues.

Moreover, idioms are often used to engage in word play. A case in point is the play on a possible literal interpretation of the non-literal idiom (e.g. *kitchen sink*, *pork barrel*, *a nail in the coffin*, *carrot and stick*), which demonstrates another type of lexical pun. Other creative variations include elliptical use, coordination with words that do not belong in the idiom, two conjoined idioms and the exchange of words between idioms in order to create the effects of alliteration or novelty. For instance, humor and irony can be seen in the headline of the New Zealand Herald *Korean Air*

VP goes nuts over nuts. In the nut rage incident, Korean Air vice president was so furious at the way a flight attendant served nuts on the first-class cabin that she ordered the taxiing aircraft to return to the gate before takeoff and kicked the cabin chief off the plane. Here, the idiom *go nuts* may have several implications such as going crazy/insane, being extremely mad/angry, showing a contempt or derision of foolishness and creating a rhyming effect along with *over nuts*.

Idioms also help to catch the eye of the reader and grab attention. They may bring the gist of a news story to the fore and draw the reader into the whole article, as shown in the feature story of the New York Times *Southwest Airlines is sitting pretty*. The idiom *sitting pretty* here illustrates the point of the topic and serves as a hook to engage the attention of readers to Southwest's seating issue. In the paragraphs that follow, the feature story reports that instead of handling the fact that they do not offer the convenience of reserved seating, Southwest Airlines used TV commercials to turn this drawback into a liberating image—free-to-sit-anywhere.

Last but not least, quite a majority of truncated idioms are used in lieu of their full forms. The elliptical forms can be explained by the economical syntax that is typical of journalistic language and especially used in news headlines. In most cases, a part of an idiom is enough for the reader with knowledge of this idiom to recall the whole (*big fish in a small pond*; the *early bird catches the worm*).

Text coverage of the most frequent opaque idioms in English news

The opaque idiom list contains 525 entries of 2 to 6 words with an accumulation of 16,334,571 individual instances and 57,171,019 running words, which makes up 0.59% of the tokens in the NOW Corpus.

At first sight, the 0.59% text coverage does not appear to be worth noticing. Compared to a plethora of idioms in a native speaker's lexicon, the frequency of a limited cohort of idioms naturally becomes relatively small. They perhaps do not deserve

classroom time due to a very small number of opportunities for meeting them. However, Martinez and Schmitt (2012) have cautioned that lexical profiling that does not take account of opaque phraseology would risk underestimating the lexical complexity of a text. Their concern is true of the present research.

To put this in perspective, several excerpts from the NOW Corpus are shown below. The idioms are underlined and in bold. They may give us a picture of the most frequent opaque idioms used in everyday news.

“I don't think there's any real **smoking gun**,” a former veteran FBI official told the New York Times, asking for anonymity because of the sensitivity surrounding the investigation.

New York Times

“I'm not seeing a big tick up in defense spending **across the board**,” Leanne Caret, who heads Boeing's defense business, told Reuters in an interview.

Reuters

“At this point, Comey is not a government employee, and he is not under the direction of the president, and so I don't see a legal or procedural bar,” Aftergood said. “He's being totally **above board** about it, and if anybody wants to say that's wrong, he is standing by his actions. If anybody wants to pursue him, he has made himself available, and that's the most anyone can ask for.”

Associated Press

While Venezuelans bitterly complain about shortages of food and medicine, few still respond to opposition calls for protests, **a far cry from** early demonstrations that saw hundreds of thousands taking to the streets.

VOA News

“Though we don't agree with Antifa's tactics and adventurism, we respect their willingness to **put** their bodies **on the line** to fight fascists,” organizer Nino Brown told NBC News.

NBC News

Among the 186 running words of news excerpts, five different idioms (15 words in total) belong to the present opaque idiom list.

Their coverage in these excerpts is 8% in tokens (=15/186). Without knowledge of these five idioms, the text coverage of the first 3,000 word families will decrease from 90% to 82%. In other words, the lexical text coverage is overestimated, unless these five opaque idioms are known. Previous research indicates that a reasonable comprehension level may be achieved at 95% lexical coverage (Laufer & Ravenhorst-Kalovski, 2010). With only 82% coverage, English-language learners may not be able to read these news passages effectively.

Native English-speaking children view a vocabulary load of two unknown words per hundred words as difficult reading (Carver, 1994). If the putative 98% lexical coverage (allowing 2% unknown words) is the optimal threshold necessary for unassisted, adequate comprehension of a text (Nation, 2006), the opaque idioms occurring in the above news excerpts (8% coverage) should not be neglected. If unfamiliar, they are very likely to impede comprehension and may cause misinterpretation due to non-literal or non-compositional meanings.

In this research, one of the selection criteria was component words within the first 3,000-word-family level and the GSL. When being integrated into the first 3,000 word families, the component words of the 525 opaque idioms (involving 554 word families, see Table 3) would constitute 18.5 percent of the first 3,000 word families and are enough to cause comprehension issues if not understood.

For a reliability check, the 525 opaque idioms were examined against 6 million words of VOA news. They were sorted based on three divisions of high-, mid- and low-frequency. Table 5 gives us a snapshot of the occurrences of the opaque idioms in the VOA and NOW Corpora. For instance, *state of the art* and *in the pipeline* rank high on the idiom list, while *run-of-the-mill* and *all the rage* are in the middle and *out of sorts* and *for a song* at the bottom. To examine their frequency in the VOA News Corpus, raw frequency was converted into occurrences per million words, since the VOA and NOW Corpora are different in size. A look at Table 5 shows that they appear in a similar pattern of frequency in the VOA News Corpus.

To confirm whether the occurrences of the opaque idioms in the NOW and the VOA Corpora have a high level of agreement, the Pearson r correlation coefficient was computed on the SPSS. The test result ($r = 0.72$, $p = .005$) verifies that there was a significant consistency in the frequency of occurrence per million words for the 525 opaque idioms appearing in both corpora.

Overall, the 525 opaque idioms accounted for 0.61% of the total words in the VOA News Corpus as opposed to its accumulated coverage of 0.59% in the NOW. The slight difference confirmed the wide use of the idioms in journalistic register. English news readers may encounter these idioms very often since they slip in and out of everyday news reports along with high-frequency words.

Table 5. High-, mid- and low-frequency idioms in NOW and VOA News

	Opaque idioms	Occurrences per million words in NOW (9.69 billion words)	Occurrences per million words in VOA (6 million words)
High-frequency	the likes of	26.30 (254,904)	28.50 (171)
	make waves	10.71 (103,778)	10.50 (63)
	state of the art on	9.12 (88,392)	7.50 (45)
	the sidelines	8.52 (82,562)	2.50 (15)
	in the pipeline	3.14 (30,482)	2.83 (17)
Mid-frequency	fit/fill the bill	1.01 (9,759)	1.00 (6)
	run-of-the-mill all	0.79 (7,717)	0.83 (5)
	the rage push the	0.76 (7,387)	0.67 (4)
	envelope	0.76 (7,354)	0.83 (5)
Low-frequency	pass the buck	0.46 (4,479)	0.33 (2)
	out of the running	0.45 (4,397)	0.33 (2)
	out of sorts	0.42 (4,117)	0.17 (1)
	for a song	0.37 (3,550)	0.17 (1)

Note: The numbers in the parentheses are raw frequency.

Conclusion

The principal concern of this study was to create a pedagogical useful list of the most frequent opaque/non-compositional idioms for English learners who use news texts as their learning material. By means of a principled set of criteria, a total of 525 idioms of 2 to 6 words were selected and they made up 0.59% of the running words in the NOW Corpus. The opaque idioms are made up of the first 3000 word families and the GSL.

Therefore, they can partially bridge the gap between the text coverage that very common words can and cannot account for in news articles to facilitate reading comprehension. The idiom list is short and may be a viable option for learners to learn within a short period of time. This research has identified the discourse functions that opaque idioms perform in English news and they may help in raising learners' awareness of the context opaque idioms occur in and how they behave in authentic texts. Despite arbitrary decisions on cut-off values in the compilation of opaque idioms, there may be some advantages to overt instruction of these idioms. As to the effectiveness of learning opaque idioms, it is worth investigation but beyond the present focus. As with other individual wordlists, it is hoped that the listing of the most frequent opaque idioms may serve as a reference for the development of journalistic English learning material.

Acknowledgements

The author is grateful to anonymous reviewers for their suggestions. This study was supported by a Grant (MOST 109-2410-H-214-00) from the Ministry of Science and Technology in Taiwan.

The Author

Wenhua Hsu is currently teaching at I-Shou University in southern Taiwan. She holds a PhD in the field of ESP from the University of Essex, UK and MBA from Kansas State University, USA. Her interest includes frequent academic/sub-technical and lay-technical words as well as multi-word units in specialized fields.

Idiom Dictionaries

Cambridge international dictionary of idioms—*itools*. (2019). Cambridge: Cambridge Dictionaries. Retrieved from <http://itools.com/tool/cambridge-international-dictionary-of-idioms>

Collins COBUILD idioms dictionary. (2012). (3rd ed.). Glasgow: HarperCollins.

- Farlex idioms and slang dictionary. (2017). Huntingdon Valley, PA: Farlex International.
- Longman idioms dictionary. (1998). Harlow, UK: Pearson Education.
- McGraw-Hill's Dictionary of American idioms and Phrasal Verbs. (2006). New York: McGraw-Hill Education.
- Oxford idioms dictionary for learners of English. (2007). Oxford: Oxford University Press.
- The free dictionary—idioms. (2019). Huntingdon Valley, PA: Farlex. Retrieved from <https://idioms.thefreedictionary.com/>
- The American Heritage Dictionary of Idioms. (2013). (2nd rev. ed.). Boston: Houghton Mifflin Harcourt.

References

- Alexander, R. J. (1987). Problems in understanding and teaching idiomaticity in English. *Anglistik und Englisch unterrichtet*, 32, 105–122.
- Anthony, L. (2014). AntWordProfiler (Version 1.4.1) [Computer Software]. Tokyo, Japan: Waseda University. Retrieved from <http://www.laurenceanthony.net/software>
- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at ...: lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25(3), 371-405. doi:10.1093/applin/25.3.371
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow, England: Pearson.
- Carver, R.P. (1994). Percentage of unknown words in text as a function of the relative difficulty of the text: Implications for instruction. *Journal of Reading Behavior*, 26, 413–437. doi:10.1080/10862969409547861
- Chen, Y., & Wang, L. (2016). Opening Pandora's box: A corpus-based study of idioms in ELT materials. *The Journal of Asia TEFL*, 13(4), 381-390. doi:10.18823/asiatefl.2016.13.4.10.381
- Cooper, T. C. (1998). Teaching idioms. *Foreign Language Annals*, 31, 255–266. doi:10.1111/j.1944-9720.1998.tb00572.x
- Fernando, C. (1996). *Idioms and idiomaticity*. Oxford: Oxford University Press.

- Fernando, C., & Flavell, R. (1981). *On idiom: Critical views and perspectives*. Exeter, UK: University of Exeter.
- Glucksberg, S. (2001). Oxford psychology series; no. 36. *Understanding figurative language: From metaphors to idioms*. New York: Oxford University Press.
- Grant, L. (2005). Frequency of 'core idioms' in the British National Corpus (BNC). *International Journal of Corpus Linguistics*, 10(4), 429-451. doi:10.1075/ijcl.10.4.03gra
- Grant, L., & Bauer, L. (2004). Criteria for re-defining idioms: Are we barking up the wrong tree? *Applied Linguistics*, 25(1), 38-61. doi:10.1093/applin/25.1.38
- Grant, L., & Nation, I.S.P. (2006). How many idioms are there in English? *International Journal of Applied Linguistics*, 151, 1-14. doi:10.2143/ITL.151.0.2015219
- Hsu, W. (2019). Voice of America (VOA) news as voluminous reading material for mid-frequency vocabulary learning. *RELC Journal*, 50(3), 408-421. doi:10.1177/0033688218764460
- Hyland, K. (2008). As can be seen: lexical bundles and disciplinary variation. *English for Specific Purposes*, 27(1), 4-21. doi:10.1016/j.esp.2007.06.001
- Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language*, 22(1), 15-30.
- Liu, D. (2003). The most frequently used spoken American English idioms: A corpus analysis and its implications. *TESOL Quarterly*, 37(4), 671-701. doi:10.2307/3588217
- Makkai, A. (1972). *Idiom structure in English*. The Hague: Mouton.
- Martinez, R., & Murphy, V. A. (2011). Effect of frequency and idiomaticity on second language reading comprehension. *TESOL Quarterly*, 45(2), 267-290. doi:10.5054/tq.2011.247708
- Martinez, R., & Schmitt, N. (2012). A phrasal expressions list. *Applied Linguistics*, 33(3), 299-320. doi:10.1093/applin/ams010
- McCarthy, M. (1998). *Spoken language and applied linguistics*. Cambridge: Cambridge University Press.
- Moon, R. (1998). *Fixed expressions and idioms in English*. Oxford: Clarendon Press.

- Nation, I.S.P. (2017). The BNC/COCA Level 6 word family lists (Version 1.0.0) [Data file]. Retrieved from <http://www.victoria.ac.nz/lals/staff/paul-nation.aspx>
- Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review*, 63(1), 59-82. doi:10.3138/cmlr.63.1.59
- Nation, I. S. P. (2005). Range program with British National Corpus list 14,000 words. Retrieved from <https://www.victoria.ac.nz/lals/about/staff/paul-nation#vocab-programs>
- Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge: Cambridge University Press.
- Simpson, R., & Mendis, D. (2003). A corpus-based study of idioms in academic speech. *TESOL Quarterly*, 37(3), 419-442. doi:10.2307/3588398
- West, M. (1953). *A general service list of English words*. London: Longman, Green, & Co.

Appendix

The 525 most frequent opaque idioms in English news

Below is a list of the 525 most frequent opaque idioms in English news, ranked in descending order of frequency. The assumption behind this frequency-based ranking is that higher-frequency idioms may be more likely to be encountered than lower-frequency idioms. In terms of a good learning return, idioms with high frequency may be worth priority attention at the initial phase of idiom learning. This frequency-based idiom list may serve as a reference and teachers may rearrange it for use in a range of ways, e.g. by themes, in alphabetical order or based on the number of words in the idiom for memory's sake. Due to limited space, five pedagogical tips are suggested here:

1. Teachers can encourage students to picture in their mind the idioms that are made up of words that create images of the expressions (*call the shots, out of pocket, put on the back burner, take a back seat*). If learners know the original context in which the idiom was used and understand the image it is based on, they would find the idiom easier to learn. For example, the idiom *in the red* alludes to 'in debt or overdrawn'. If learners know the bookkeeping practice of writing money outflow in red ink and money inflow in black, it may be effortless for them to get the picture of losing money and turning a profit for the idioms *in the red* and *in the black*.
2. Learning idioms by themes may be helpful. By grouping idioms according to the domain they originate from (e.g. warfare, sports, theater, cuisine), it may be easier for learners to recognize the connections and remember their meanings (in horse-racing, *jump the gun, neck and neck, win hands down*; in card games, *across the board, upper hand*).
3. Although this opaque idiom list was derived from the largest news corpus, itemized idioms are still not enough for English learners. As with the learning of individual words, idioms should be learned in context rather than as disembodied items. English teachers can raise their students' awareness of how idioms behave in authentic discourse. Through access to free online corpora or web concordancers (e.g. BYU Corpora, Compleat Lexical Tutor), abundant contextual examples can be obtained from concordance lines for classroom exercises (e.g. idiom cloze with glosses provided).
4. Learners should be alerted to the fact that idioms in authentic discourse do not always occur in their canonical forms. Variations of idioms, creative blending and unpredictable uses often occur.
5. Learners may start with English idioms that have similar meaning in their native language (L1). L1 idioms are actually a good support to learning English idioms with identical meanings. With knowledge of mother-tongue idioms, learners may find it less difficult when learning new English idioms (e.g. *break even* in both Chinese and English, meaning profits and costs in equal amounts).

The 525 opaque idioms are listed in decreasing order of frequency and divided into three ranks for space sake.

Rank 1		Rank 2		Rank 3	
Opaque idioms	Frequency	Opaque idioms	Frequency	Opaque idioms	Frequency
the likes of	254,904	in the bag	10,347	be that as it may	5,243
to death	198,654	a blind spot	10,292	put one's best foot forward	5,242
in (the) light of	136,520	dark horse	10,264	in pole position	5,238
make waves	103,778	bad blood	10,251	smoking gun	5,225
go viral	95,811	a labor of love	10,161	hard and fast	5,153
bottom line	91,042	take one's cue from	10,160	Pandora's box	5,147
in the long run/term	88,576	seal/stamp of approval	10,101	with flying colors	5,100
state-of-the-art; state of the art	88,392	clean sweep	10,091	Big Brother	5,070
on the sidelines	82,562	arm's length	9,989	hands down	5,066
all out; all-out	76,996	a breath of fresh air	9,928	put ~ on the map	5,056
fall short (of)	73,897	on a par (with)	9,886	not pull one's punches; pull no punches	5,005
on and off; off and on	70,576	wrap/get one's head around/round	9,827	a call to arms	4,973
behind the scenes	66,109	iron hand/fist (in a velvet glove)	9,803	thick and thin	4,971
big deal	63,961	foot the bill	9,773	against the grain	4,910
play the game	54,922	fit/fill the bill	9,759	make a quick/fast buck	4,872
in one's (hip) pocket	52,567	fall on deaf ears	9,740	of the old school	4,867
under fire	49,715	on a roll	9,645	an about-face	4,823
across the board	48,834	have a go at	9,613	cold feet	4,814
around the corner	48,006	on the fly	9,606	make up for lost time	4,781
have been around	46,471	without fear or favor	9,579	hold water	4,759
in the short run/term	46,317	a big shot	9,442	black sheep	4,731
follow suit	46,304	[get/set/start/keep] the ball rolling	9,367	hard feelings	4,714

Rank 1		Rank 2		Rank 3	
Opaque idioms	Frequency	Opaque idioms	Frequency	Opaque idioms	Frequency
[get/have] the [better/best] of	45,136	[set/cast/carved] in stone	9,365	reinvent the wheel	4,693
a rolling stone	42,190	back to the drawing board	9,339	face the music	4,677
in the works	41,883	clear the air	9,296	an iron curtain	4,672
over the top	41,748	a grey/gray area	9,275	jump the gun	4,606
on the horizon	40,305	last but not least	9,242	thick and fast	4,606
be all in	40,027	flat out	9,233	of one's own accord	4,594
from scratch	38,814	find one's feet/feel one's legs	9,166	been there, done that	4,577
walk of life	38,630	an [ace/card/trick] up one's sleeve	9,156	under a cloud	4,576
down the line	37,136	in black and white	9,149	fourth estate	4,542
to name (but) a few	36,560	make the best of it	9,139	sitting pretty	4,515
on the back foot	34,847	die hard	9,082	pass the buck	4,479
in full swing	34,051	a second thought	9,055	give and take	4,463
in the red	33,935	a stone's throw	9,013	a long arm	4,414
on/in the cards	33,817	have one's work cut out	8,993	(right) on cue	4,403
on the ball	33,796	know better than	8,952	out of the running	4,397
keep/hold ~ at bay	33,326	get one's act together	8,945	a can of worms	4,382
under one's belt	31,975	the writing on the wall; the writing is on the wall	8,928	in the final analysis	4,381
catch fire	31,204	fly in the face of	8,895	under one's (very) nose	4,352
in the pipeline	30,482	pride of place	8,848	toe the line	4,312
make no mistake	29,991	rub shoulders with	8,844	rock the boat	4,306
up for grabs	29,479	take a back seat	8,778	act of God/Nature	4,285
red tape	28,599	sit on the fence	8,774	get the hang of	4,268
all over the place	28,547	ahead of one's time	8,667	true blue	4,228

Rank 1		Rank 2		Rank 3	
Opaque idioms	Frequency	Opaque idioms	Frequency	Opaque idioms	Frequency
in one's (own) backyard/ back yard	28,133	a rainy day	8,571	pour/throw cold water on	4,218
all eyes	27,388	the best of both worlds	8,535	give something a miss	4,174
blow the whistle; whistle blower; whistle blowing	26,996	throw in the towel	8,467	banana republic	4,126
down to earth	25,800	on one's plate	8,425	from the word go	4,118
think twice	24,752	on the shelf	8,425	out of sorts	4,117
a house of cards	24,592	neck and neck	8,423	break the ice	4,109
make (both) ends meet	24,562	with/without a grain/pinch of salt	8,369	raw/rough deal	4,065
behind closed doors	24,222	come to a head; bring ~ to a head	8,333	anybody's guess	3,995
for the record	24,038	shot in the arm	8,301	a bad/rotten apple	3,886
left and right; right and left	23,832	out of order	8,289	not mince words	3,849
no brainer	23,659	out of the wood(s)	8,281	a toe/foot in the door	3,838
green light	23,412	a double-edged sword/weapon	8,233	music to one's ears	3,781
once and for all	23,295	tall order	8,230	a stab in the back	3,777
witch hunt	22,958	big fish (in a small pond)	8,184	the party's over	3,761
set the stage for	22,937	set the pace	8,093	see eye to eye	3,753
early bird (catches the worm)	22,734	out of line	8,051	have/keep one's feet on the ground	3,735
thumbs up	22,659	drag one's feet or heels/ foot dragging	8,025	pull one's weight	3,727
a far cry from	22,552	read between the lines	8,024	miss the boat/bus	3,725
turn a blind eye	22,534	in the flesh (=in person)	8,005	take someone for a ride	3,718
hold one's own	22,278	give someone a break	7,999	a slap on the wrist	3,718

Rank 1		Rank 2		Rank 3	
Opaque idioms	Frequency	Opaque idioms	Frequency	Opaque idioms	Frequency
take stock of	22,130	ride high	7,904	after one's own heart	3,660
in the black	21,675	rear its (ugly) head	7,875	AWOL (absent without leave)	3,660
out of pocket	21,614	below par	7,771	come/go with the territory	3,652
around/round the clock; the clock round	21,427	run of the mill	7,717	take the cake/biscuit	3,648
upper hand	21,068	go the (full) distance	7,700	smoke and mirrors	3,635
by and large	21,048	deep pockets	7,692	bury one's head in the sand	3,628
once in a while	20,495	go overboard	7,669	ring a bell	3,626
A1	20,446	in deep water/waters	7,569	drop dead	3,622
raise one's eyebrows	20,328	call it a day	7,532	bite the bullet	3,606
on the same page	20,223	under the skin	7,526	out of the closet	3,600
at par	19,780	before you know it	7,518	behind the curtain	3,587
(think) outside the box	19,432	down the drain/ tubes	7,491	cover one's tracks	3,585
(up) in the air	19,113	tooth and nail	7,417	have a lock on	3,585
have a go (at)	19,111	in the same/next breath	7,411	take someone's breath away	3,582
be/hang in the balance	18,717	all the rage	7,387	a sacred cow	3,569
by/through the back door	18,529	white elephant	7,376	flesh and blood	3,562
a silver lining	18,418	flex one's muscles; muscle flexing/muscle-flexing	7,367	sit on one's hands	3,554
fair play	18,200	play ball (with)	7,362	bend/lean over backward(s) (to do something)	3,553
on one's toes	17,941	push the envelope	7,354	for a song	3,550

Rank 1		Rank 2		Rank 3	
Opaque idioms	Frequency	Opaque idioms	Frequency	Opaque idioms	Frequency
fingers crossed; cross one's fingers	17,823	bells and whistles	7,320	fat cat	3,534
have what it takes	17,667	come/get to grips with	7,272	moment of truth	3,533
be taken aback/ take someone aback	17,481	mean business	7,251	In the heat of the moment	3,530
in aid of	17,378	food for thought	7,219	lame duck	3,519
above and beyond	16,858	Achilles heel	7,211	a chip on one's shoulder	3,518
make (it) big	16,604	break the bank	7,172	a cut above	3,517
set one's sights on	16,319	song and dance	7,121	good riddance	3,509
as and when	16,249	drop the ball	7,076	fill someone's shoes/boots	3,492
out and about	16,225	pull out all the stops	7,034	a bit much	3,489
in the market for	16,068	stay the course/distance	7,027	on the mend	3,462
in a nutshell	16,032	turn the corner	6,988	a halfway house	3,448
make or break	16,017	break new ground	6,984	the eye of the storm	3,439
out of the blue	15,922	no mean	6,971	out of thin air	3,429
off the mark; wide of the mark	15,802	come full circle	6,956	rewrite history	3,403
on the house	15,759	cut one's teeth on/in/at	6,932	divide and rule	3,396
steal the show/scene	15,708	tighten one's belt; belt-tightening	6,907	red-light district	3,390
up to speed	15,625	(right) on the money	6,906	shout something from the rooftops	3,389
sweep/brush ~ under the rug/carpet	15,144	olive branch	6,899	(draw) a line in the sand	3,387
bear the brunt of	14,997	keep/leave/have one's options open	6,880	worth one's salt	3,372
for good measure	14,914	last laugh	6,863	crash and burn	3,368
last ditch	14,893	go downhill	6,828	go easy on	3,322

Rank 1		Rank 2		Rank 3	
Opaque idioms	Frequency	Opaque idioms	Frequency	Opaque idioms	Frequency
to the letter	14,734	pull the strings	6,800	on cloud nine	3,321
heart and soul	14,654	jump ship	6,774	take/call someone to task	3,317
part and parcel	14,650	take the law into one's (own) hands	6,720	a means to an end	3,307
hit the road	14,468	down to the wire	6,616	down to the ground	3,299
to boot	14,193	in the same boat	6,612	top dollar	3,292
point the finger; finger pointing	14,084	hot on the heels of	6,521	to a tee / T	3,279
at the eleventh hour	14,005	cut corners	6,499	from A to Z	3,269
come clean	13,971	claim to fame	6,482	head over heels	3,235
make do	13,794	a nail in the coffin	6,469	line one's pocket/pockets	3,190
rule of thumb	13,753	off the wall; off-the-wall	6,456	sour grapes	3,163
a long shot	13,670	go through the roof; hit the roof	6,448	pick up steam	3,152
come of age	13,610	buy time	6,374	gone with the wind	3,134
make the grade	13,527	[for/to] all intents and purposes	6,373	a thorn in someone's side/flesh	3,126
serve notice	13,490	tip the scales	6,373	the wheel of Fortune	3,097
hit the ground running	13,482	bang for one's buck	6,371	out on a limb	3,072
under the banner of	13,476	the middle of nowhere	6,326	pack a punch	3,050
sweet spot	13,372	play (right) into someone's hands	6,314	seal someone's fate	3,034
keep something under wraps	13,338	out of your mind	6,296	as a matter of course	2,985
off the hook	13,285	lend (someone) a hand	6,291	add fuel to the fire/flames	2,968
beg the question	13,248	feel the pinch	6,288	between a rock and a hard place	2,961

Rank 1		Rank 2		Rank 3	
Opaque idioms	Frequency	Opaque idioms	Frequency	Opaque idioms	Frequency
in the dock	13,094	stick to one's guns	6,202	faint of heart; a faint heart	2,942
good Samaritan	12,891	under the counter/table	6,166	in stitches	2,939
hit/make the headlines	12,704	a safe bet	6,145	(from) hand to mouth	2,937
rain or shine	12,687	sky is the limit; sky's the limit	6,120	drop/lower one's guard; let down one's guard	2,933
get (something) off the ground	12,680	in spades	6,111	brownie points	2,928
in the wind	12,678	catch-22	6,097	vote with one's feet	2,925
pay lip service	12,675	top dog	6,074	poor man's	2,917
over the moon	12,671	to the bone	6,054	dead in the water	2,911
tongue-in-cheek	12,634	off the record	6,046	bite the dust	2,891
leave no stone unturned	12,566	off the cuff	6,035	have it both ways	2,891
pull the plug	12,421	a second string to one's bow	5,996	a necessary evil	2,880
on [that/this] [note/score]	12,402	anything goes	5,969	a purple patch/passage	2,867
under par	12,386	[learn/know/s how] the ropes	5,938	hot potato	2,865
bear fruit	12,313	spill the beans	5,928	hit the jackpot	2,859
pull the plug (on)	12,234	lose steam; run out of steam/gas	5,924	over the hill	2,857
in the driver's/driving seat	12,146	a hidden agenda	5,914	two-way street	2,856
bread and butter	12,055	carry the ball	5,902	put your foot in it	2,830
in broad daylight	11,981	(not) up to scratch	5,896	the bare bones	2,829
second to none	11,956	out of the picture	5,871	cold shoulder	2,821
pork barrel	11,939	kitchen sink	5,870	couch potato	2,800
in the loop	11,871	back to square one	5,827	call a spade a spade	2,800
hit/strike home	11,864	a pat on the back	5,764	press the button	2,796
put/set the	11,851	par for the	5,751	rise from the	2,795

Rank 1		Rank 2		Rank 3	
Opaque idioms	Frequency	Opaque idioms	Frequency	Opaque idioms	Frequency
record straight		course		ashes	
do the trick	11,757	on top of the world	5,742	acid test	2,791
against all (the) odds	11,691	in no uncertain terms	5,725	close ranks	2,791
the last/final straw	11,647	someone's cup of tea	5,724	on the square	2,779
out of the race	11,538	a drop in the bucket/ocean	5,669	front burner	2,779
few and far between	11,518	far and away	5,651	out of the loop	2,775
call the shots	11,489	blow/lose one's cool	5,611	better half	2,773
a silver/magic bullet	11,430	above board	5,608	the other side of the coin	2,734
(wait) in the wings	11,380	in short order; in ~ short order	5,591	dead and buried	2,733
rise to the occasion	11,313	the jury is (still) out	5,562	make no bones about	2,712
lock horns	11,196	mover and shaker	5,559	drop names	2,698
give it a shot/try	11,131	get under one's skin	5,520	by dint of	2,687
as it were	10,979	weather the storm	5,511	for my money	2,672
(drive) under the influence	10,950	beg to differ	5,502	beyond the pale	2,652
ahead of the game	10,787	the name of the game	5,496	flash in the pan	2,631
benefit of the doubt	10,768	in good time	5,460	at the drop of a hat	2,613
the powers that be	10,751	err on the side of	5,446	on the mark	2,596
on the rocks	10,722	put paid to	5,433	neither here nor there	2,587
put/lay ~ on the line	10,650	go places	5,430	cry wolf	2,564
the beginning of the end	10,633	back burner	5,375	speak the same language	2,549
turn the tables; the tables are turned	10,622	on the ropes	5,373	crocodile tears	2,536
icing on the cake	10,608	a class act	5,368	sweep something under the carpet/rug	2,529

Rank 1		Rank 2		Rank 3	
Opaque idioms	Frequency	Opaque idioms	Frequency	Opaque idioms	Frequency
up in arms	10,561	up to scratch	5,366	(have) an ax/axe to grind	2,527
the order of the day	10,547	thumbs down	5,353	cold turkey	2,521
under the hood	10,544	right/straight off the bat	5,316	up/down one's alley; right up one's street	2,506
by leaps and bounds	10,520	in the hole	5,299	be/feel (all) at sea	2,504
in hot water	10,442	paper over (the cracks)	5,294	a piece of cake	2,503
elephant in the room/corner	10,390	call someone names	5,277	touch base	2,500