Idioms or Open Choice? A Corpus Based Analysis

Kaitlyn Alayne VanWagoner
Brigham Young University

Follow this and additional works at: https://scholarsarchive.byu.edu/etd

BYU ScholarsArchive Citation
VanWagoner, Kaitlyn Alayne, "Idioms or Open Choice? A Corpus Based Analysis" (2017). All Theses and Dissertations. 7291.
https://scholarsarchive.byu.edu/etd/7291
Idioms or Open Choice? A Corpus Based Analysis

Kaitlyn Alayne VanWagoner

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Arts

Dee Gardner, Chair
Mark E. Davies
Don W. Chapman

Department of Linguistics and English Language
Brigham Young University

Copyright © 2017 Kaitlyn VanWagoner
All Rights Reserved
ABSTRACT

Idioms or Open Choice? A Corpus Based Analysis

Kaitlyn Alayne VanWagoner
Department of Linguistics and English Language, BYU
Master of Arts

The ambiguous nature of idioms has been a persistent challenge for English language learners and researches alike. Problematic issues include identifying which idioms are most pertinent for study, and the question of how frequently idiom forms found in dictionaries and other canonized resources actually function as idioms in real language use. This study differentiates between idiom forms used idiomatically (idiom-principle) versus literally (open-choice principle), and provides quantitative data to assess this difference. The data was obtained through a corpus analysis of 1,000 randomly-selected idioms in 10,000 randomly-selected contexts (10 contexts per idiom), and revealed that the majority of idiom forms were indeed functioning idiomatically in the contexts analyzed, but there were also notable exceptions. The findings are used to support the general notion in the literature that idioms represent a single lexical choice for language users, and the researcher proposes several extensions of the findings for the teaching and researching of idioms.

Keywords: idioms, idiom-principle, open-choice, corpus analysis
ACKNOWLEDGEMENTS

I would like to express my gratitude to my committee members for their dedication and time. Also, my profound thanks to my thesis advisor Dr. Dee Gardner of the Linguistics and English Language Department at Brigham Young University for his unfailing patience and wonderful direction and tutelage on the present document as well as in my journey as a scholar and as an individual. I must also thank my parents for their love, support, and the many opportunities they provided to me, without which this accomplishment would not have been possible. Thank you.
Table of Contents

TITLE PAGE ...........................................................................................................................................i
ABSTRACT ...............................................................................................................................................ii
ACKNOWLEDGEMENTS .......................................................................................................................iii
TABLE OF CONTENTS ............................................................................................................................iv
LIST OF FIGURES .......................................................................................................................................v
LIST OF TABLES .........................................................................................................................................vi
Chapter 1: Introduction .............................................................................................................................1
Chapter 2: Review of Literature ................................................................................................................3
  Defining Idiom .....................................................................................................................................3
  Idiomaticity .........................................................................................................................................5
  Idiom vs. Open-Choice .......................................................................................................................8
  Research Aims ..................................................................................................................................10
Chapter 3: Methodology .........................................................................................................................12
  Idiom Definition for this Study ..........................................................................................................12
  Idiom Source ....................................................................................................................................12
  Identifying a Sample Size of Idioms .................................................................................................13
  Corpus Selection ...............................................................................................................................13
  NOW Corpus ....................................................................................................................................13
  Context Generation ...........................................................................................................................14
  Analysis of Idioms ............................................................................................................................18
    Interrater Discrepancy .....................................................................................................................20
Chapter 4: Results and Discussion .......................................................................................................23
  Results ...............................................................................................................................................23
    Idiom Principle (True Idioms) .........................................................................................................25
      Open-Choice Principle (False Idioms) .........................................................................................26
    Idioms as Proper Nouns ..................................................................................................................28
  Discussion .........................................................................................................................................29
Chapter 5: Conclusion ............................................................................................................................31
  Limitations .........................................................................................................................................32
  Extensions ..........................................................................................................................................33
  Future Research ...............................................................................................................................34
References ...............................................................................................................................................35
Appendix A ............................................................................................................................................38
Appendix B ............................................................................................................................................47
List of Figures

Figure 1. Idiom entry .................................................................................................................... 14
Figure 2. Idiom Contexts ............................................................................................................. 15
Figure 3. Excel spreadsheet .......................................................................................................... 15
Figure 4. Idiom wildcard entry ..................................................................................................... 16
Figure 6. Idiom contexts ............................................................................................................... 17
Figure 7. If the first idiom form has less than 10 contexts, the second form is also collected. .... 18
Figure 8. Dichotomous rating in Excel spreadsheet ..................................................................... 19
Figure 9. Sample of first and second rater system ........................................................................ 19
Figure 10. Degree of Idiomaticity (Note: rounding caused the total percentage to be less than 100) ............................................................................................................................................... 24
List of Tables

Table 1. Interrater Discrepancy Examples................................................................. 22
Table 2. Idiomatic language examples...................................................................... 25
Table 3. Examples of False Idioms as Open Choice ............................................... 27
Table 4. Idioms functioning as Proper Nouns .......................................................... 28
Chapter 1: Introduction

Idioms are a perplexing linguistic challenge, often causing comprehension difficulties, especially for second language (L2) learners, unless the meaning has been explicitly taught. Grant and Nation (2006) point out that without direct instruction, most English language learners will understand idioms literally, rather than figuratively, and be confused. Research supports the theory that English language learners are more likely to interpret idioms literally than native speakers, who are more familiar with the phrasal structures and culture of English and thus are able to search for meaning in the language. However, even native speakers will misunderstand an idiom with which they are unfamiliar. The necessity of idioms in language is illustrated by Cooper’s (1998) estimate that around 980 idiomatic types are used by individuals daily, accumulating to approximately 20 million in a lifetime. The complexity of idiom usage is clearly established. However, language users and instructors may be overwhelmed by the challenge of identifying common idioms for instruction.

In research, idioms have proven difficult to define, identify, and quantify. While there are various studies regarding idiom acquisition and comprehension, as well as the form and variety of idioms in language, there is a need for quantification of how these idioms are functioning in authentic language. In order to identify idioms most relevant for study, it is necessary to know which idiom forms are most reliably functioning as idioms. The challenge is to differentiate between when a single idiom, such as wrap it up, is being used figuratively, rather than literally. For example, is wrap it up being used, metaphorically, to indicate the conclusion of an activity? Or is it requesting, literally, that an item be wrapped in paper? Modern electronic corpora have provided a means to investigate the usage and frequency of multiword items, such as idioms, with a quantitative approach impossible prior to this technological advancement, when linguists
had to rely primarily on intuition and hand calculations to conduct such studies. Indeed, electronic corpora allow large quantities of text to be objectively processed in quantities that were once considered impossible, bringing new issues to light and allowing multiword items (idioms, phrasal verbs, stock phrases, etc.) to become significant subjects in English language research and in English language education.

John Sinclair (1991) embraced corpus research and utilized it in his theorization of the construct of idiom. He argues that multiword items do not necessarily occur at random in the language and that they do, in fact, function as semi-preconstructed phrases, constituting single choices. This model of interpretation is known as the “idiom principle” and stands in opposition to what Sinclair dubbed the “open-choice principle,” which prescribes that grammatical language units allow users to select from a range of lexical (vocabulary) choices. Sinclair’s concept of the idiom principle has been widely embraced by linguists examining idioms. Researchers such as Grant & Nation (2006) and Levorato, Roch & Nesi (2007) have asked how often language speakers can count on an item identified as an idiom being used idiomatically rather than literally, but no quantitative study has successfully answered this query. The present study will attempt to do so by analyzing 1,000 randomly selected idioms in 10,000 randomly selected contexts to determine the extent to which supposed idioms are actually functioning idiomatically (idiom principle) versus literally (open choice principle).
Chapter 2: Review of Literature

Defining Idiom

Linguists have grappled with the creation of a definition for the term idiom, and criteria for definition varies. Some linguists use “idiom” interchangeably with terms such as “slang,” “jargon,” “colloquialism,” “saying,” and “proverb.” The designated length of an idiom varies as well—from Hockett’s (1965) theory that every individual morpheme may act as an idiom (e.g., the suffix /-ij/ in Chinese, pavement, New York), to Makkai’s (1972), which stipulates that an idiom must consist of at least two free morphemes (e.g., hot dog, red herring, brown betty). These contrast with those who focus on the usage of the form, rather than the form itself. Sinclair’s (1991) “idiom principle” argues that idioms are “semi-preconstructed” phrases that constitute a single choice for the language user, and, consequently, cannot be broken down even when the phrase appears to be composed of analyzable segments (p. 110).

In 2000, researchers Erman and Warren (2000) built upon Sinclair’s original definition expanding it to include conventionalized phrases and language, “chunking,” as essential for language fluency (p. 30). Examples of “chunking” include “not bad” (meaning good) and “I’m afraid” (often used to deliver bad news). These lexical phrases may vary in length and literalness, but are limited by what Warren and Britt call “restricted exchangeability.” Restricted exchangeability describes prefabricated lexical phrases whose meaning or function would be changed if a single part of the phrase were replaced by a synonymous item. For example, “good friends in they are good friends cannot be changed into nice friends without losing the implication of reciprocity” (2000, p. 32). The meanings of the whole cannot be deduced simply from knowing the meanings of the parts. Further distinction was given by the research of Nunberg, Sag, and Wasow (1994) who showed that idioms also operate within a spectrum of
idiomaticity, ranging from relatively transparent (e.g., *industrial revolution, passing lane, gain the advantage*) to opaquely idiomatic (e.g., *pull strings, pull your leg, spill the beans*). They refer to the elements Skoufaki (2009) identifies “as informality, figuration, and transparency” as key elements for the classification of idioms (p. 20). Overall, these definitions are varied and potentially ambiguous.

The challenges of idiom transparency are confronted by Fernando (1996) who developed a continuum to measure varying degrees of literalness: non-literal (e.g., *smell a rat, chip on your shoulder, break a leg*), semi-literal (e.g., *catch fire, fine tune, in the spotlight*), and literal (e.g., *tall, dark and handsome, actions speak louder than words*) (p. 32). Gibbs (1991) refers to these varying degrees of literalness as “nondecomposable” (e.g., *kick the bucket*), “abnormally decomposable” (e.g., *carry a torch*), and “normally decomposable” (e.g., *lay down the law*) (pp. 613-614). Grant and Bauer (2004) commiserate with the general verdict that idioms are not well defined and identify them as a kind of multi-word unit (MWU), and then proceed to define a MWU as a “fixed and recurrent pattern of lexical material sanctioned by usage” (p. 38). Their interpretation acknowledges that context is generally essential to identifying the use of an idiom—for example, context is essential to identify if the phrase *kick the bucket* is being used in a literal sense to describe someone physically swinging their leg to hit a bucket (*The boy kicked the bucket of apples*), or in a metaphorical sense to express that someone died (*The old dog finally kicked the bucket*). This nonliteral element of MWU is embraced by Cooper (1999) who defines an idiom as “an expression whose meaning cannot always be readily derived from the usual meaning of its constituent elements” (p. 233). For example, the idioms *kick the bucket, bite the dust, and pushing up daisies* all mean to die, but the actual word *dead* or *death* is not found in the individual words of any of these MWUs. A similar approach is taken by Siyanova-Chantura
and Martinez (2015) who use the term “multi-word expressions (MWEs)” to refer to these recurring word combinations that are semi-fixed (contain little variation in form) and include, not only idioms, but proverbs (better late than never), multi-word verbs (put up with), binomials (black and white), collocations (strong tea), speech formulae (What's up), lexical bundles (in the middle of) and more.

Disparities in idiom definitions occur because “idiom is an ambiguous term, used in conflicting ways” (Moon, 1998, p.3). This has led to varied conclusions as to the number of actual idioms in the English language. Estimates vary as widely as a mere 103 (Grant & Nation, 2006) to well over 10,000 (Brenner, 2003). Further disparity is found in the variations between a canonized idiom and how the idiom is used in fluent speech (Simpson & Mendis, 2003). When creating a list of core English Idioms, Lynn (2016) points out that “it is unavoidable that varying criteria and methods will be generated based on the scope and context of a given study” (p. 4) because of the complexity of idioms and the diversity of methodological approaches. Lynn adopts Liu’s (2008) definition of idiom: “[a] multiword expression that [is] invariant or variance-restricted in structure and often (not always) non-or semi-literal in meaning” (p. 15-16). Lynn’s liberal definition allowed him to consider all idiom entries from several established idiom dictionaries for his research, while phrasal verb dictionaries were excluded based on the conclusion drawn by Gardner and Davies (2007) that phrasal verbs deserve separate attention.

**Idiomaticity**

The great diversity in idioms mandates consideration of the relative transparency of these lexical items. Idioms have come about as the metaphorical origin of a phrase has been lost over time and “left idioms as frozen expressions” (Keysar & Bly, 1995, p. 92). Nunberg (1977) identified a native speaker’s, or advanced language learner’s, ability to utilize linguistic
knowledge to comprehend an idiom’s decomposability, a concept that divided idioms into decomposable idioms and non-decomposable idioms. Another idiom scholar, Skoufaki (2009), embraced Nunberg’s proposition and defined these terms: 1. decomposable idioms: when “a link between an idiom’s meaning and its form can be found by a native speaker” and 2. non-decomposable idioms: when “no such link can be found” (p. 23). These definitions lead into Skoufaki’s observation of Nunberg’s definition of transparency: idiom comprehension will be considerably less for idioms that are less transparent, or non-decomposable, than for idioms that are highly transparent, or decomposable. The term idiomaticity refers to this spectrum of idiom transparency. Fernando (1996) points out that “all idioms…show idiomaticity…. However, all word combinations showing idiomaticity…are not idioms”; she identifies three categories of idiomaticity: non-literal (e.g., foxglove, eavesdrop, pick-me-up), semi-literal (e.g., baby-sitter, sickroom), and literal (e.g., mother-in-law). The more “pure” an idiomatic expression is, the more linguistic knowledge is required (p. 30).

Moon (1998) synthesized Fernando’s ideas into three classifications of idioms: transparent idioms, semi-transparent idioms, and opaque idioms. Transparent idioms are those that have been “institutionalized” but have a meaning more easily deduced by highly skilled language users based on their linguistic experience, thus giving native speakers an advantage. Examples include alarm bells ring, behind someone’s back, and pack one’s bags. Conversely, semi-transparent idioms only contain partial clues that can allow their meaning to be deciphered, although there may be multiple interpretations. Examples of idioms that contain enough literalness to help convey their meaning include grasp the nettle, on an even keel, the pecking order. Last of all, opaque metaphors (idioms) are those where comprehension is not aided by any part of the idiom itself and therefore impossible to comprehend without direct knowledge of
the meaning of the metaphor. (e.g., *bite the bullet, kick the bucket, over the moon, red herring*).

While the different classifications of idioms discussed above may influence comprehension in different ways depending on whether they are opaque, semitransparent, etc., Liu (2003) points out that all idiomatic expressions create difficulties for ELLs. From the perspective of the current study, even those idioms deemed “transparent” could cause problems for language users if they appear frequently in open-choice language usage (e.g., “The alarm bells ring every time that door is opened”) as they do in idiomatic contexts (e.g., “The alarm bells ring every time I’m near that person.”). To be clear, the current study focuses on this very issue—namely, that idiom forms found in canonized idiom dictionaries and other accepted resources can appear in literal (Open Choice Principle) contexts in authentic language (the boy *kicked the bucket* of apples) or in nonliteral contexts (the old dog finally *kicked the bucket*)—the Idiom Principle.

It is important to emphasize that all studies reviewed thus far deal with the internal properties of idioms (the degree to which parts of idioms contribute directly or literally to the meanings of those idioms), whereas the current study is more concerned with how potential idioms function in actual contexts (i.e., figuratively or literally). While the first issue (the measure of internal idiomaticity) has been heavily researched, the second issue (how frequently idiom structures function as such) has very little information about it, but it has important ramifications for learning, teaching, and researching idioms. The two issues are not mutually exclusive, however, because the internal properties of idioms (relative literalness) can certainly interact with the intended meanings of larger contexts to create problems for researchers and language learners attempting to judge whether certain phrases are idioms (Idiom Principle) or not (Open Choice). For purposes of clarity and discussion, idiom forms functioning as true idioms in
actual context (Idiom Principle) will hereafter be referred to as *true idioms*, and idiom forms functioning literally in actual contexts (Open Choice) will hereafter be referred to as *false idioms*.

**Idiom vs. Open-Choice**

Because the distinction between idiom and open choice is at the heart of the current study, a more in-depth discussion seems warranted. The open-choice principle is described by Sinclair as follows:

This is a way of seeing language text as a result of a very large number of complex choices. At each point where a unit is completed (a word or a phrase or a clause), a large range of choice opens up and the only restraint is grammaticalness. This is probably the normal way of seeing and describing language. It is often called a ‘slot-and-filler’ model, envisaging texts as a series of slots which have to be filled from a lexicon which satisfies local restraints. At each slot, virtually any word can occur. Since language is believed to operate simultaneously on several levels, there is a very complex pattern of choices in progress at any moment, but the underlying principle is simple enough. (p. 109).

Sinclair offers the following examples of open choice language use contrasted with idioms: *run a mile* (idiom: “Any normal Londoner would *run a mile* rather than lunch in a Westminster pub.” ; open choice: “How fast can he *run a mile*?” ), *kick up* (idiom: “Taste it, and, if desired, *kick up* its taste a little more by whisking a bit more of the flavourings…in.”; open choice: “Slade’s brave and brilliantly-judged penalty *kick up* the touchline.”), and *stick out* (idiom: “…to find the activity and users that *stick out* as abnormal.”; open choice: “…’Klitschoko pulled a USB *stick out* of his pocket.”)
The idiom principle posits that words do not always occur as randomly as the open-choice principle postulates. Rather, words frequently co-occur, and normal text/speech is rarely produced using the open-choice principle alone:

The principle of idiom is that a language user has available to him or her a large number of semi-preconstructed phrases that constitute single-choices, even though they might appear to be analyzable into segments…. At its simplest, the principle of idiom can be seen in the apparently simultaneous choice of two words for example, of course. This phrase operates effectively as a single word, and the word space, which is structurally bogus, may disappear in time, as we see in maybe, anyway, and another. (p. 110)

The idiom principle imposes restrictions upon written and oral text and creates semantic predictability that depends on topic, situation and context. Moon (1998) explains it in these terms: “A single choice in one slot may be made which dictates which elements will fill the next slot or slots, and prevents the exercise of free choice” (p. 29). Siyanova-Chantura and Martinez (2015) analyze recurring lexical combinations, which they identify as multi-word expressions (MWEs) and refer to Sinclair’s idiom principle saying that these “combinations of words … co-occur more often than would be expected by chance alone” and, consequently, exemplify the idiom principle (e.g., spill the beans, strong tea, black and white). A key characteristic of the idiom principle, as compared to the open-choice principle, is restricted exchangeability, meaning “that at least one member of the prefab cannot be replaced by a synonymous item without causing a change of meaning or function and/or idiomaticity” (Erman & Warren, 2000, p. 32). Likewise, Liu (2008) differentiates between prefabricated phrases, which allow no variation in structure, and semi-prefabricated phrases, which permit some variance in structure, while acknowledging that both categories are, nevertheless, examples of the idiom principle that
constitutes a single choice, at a phrasal level, for the language user. While the idiom principle includes such aspects as collocations, binomials, phrasal verbs, stock phrases, proverbs, etc., as well as idioms, this study will only examine canonical idioms as they seem to be the most problematic of all such language structures.

To reiterate, the current study is not concerned with relative idiomaticity of the supposed idioms, because relative transparency can pose different comprehension challenges for language learners and native speakers alike. Rather, the present concern is the idea that sometimes these structures identified as idioms can appear with literal meanings in authentic contexts. In fact, even the most opaque of idioms can be realized literally (e.g., “the painter kicked the bucket off the step as he descended.”), and therefore not be idioms at all. How often this occurs is the central question of the current study. To the researcher’s knowledge this is the only large scale study of its kind.

Research Aims

The review of literature regarding the definition of idioms and the measurement of idiomaticity shows the complexity of defining “idiom.” The literature also reveals a gap in studies regarding idiomaticity. While much has been written about relative idiomaticity, and its associated issues, no large-scale work has been done to measure how often supposed idioms actually function idiomatally in authentic contexts (Idiom Principle) and how often they function simply as literal usages of language that just happen to have the same form as a recognized idiom (Open Choice). With this in mind, the current study has the following research aims:

1. To provide quantitative data regarding how often supposed idioms (as identified in idiom dictionaries and other resources) function idiomatically versus open-choice.
2. To determine, if possible and applicable, any qualitative differences in idioms based on the data obtained from #1.
Chapter 3: Methodology

The purpose of this study is to analyze a sample of documented idioms to discover how often they actually function idiomatically in real contexts. To accomplish this, the following steps will be taken:

1. Select a large and reliable list of English idioms.
2. Identify a reliable sample size of idioms to test from the list obtained in Step 1.
3. Collect randomly selected contexts (containing the idioms from Step 2) from a modern and reliable language corpus.
4. Analyze the idiom forms (obtained in Step 2) in their contexts (obtained in Step 3) to determine if they are functioning idiomatically or not.

Idiom Definition for this Study

The present study considers idioms from the perspective adopted by Lynn (2016)—i.e., “if an idiom occurs in an idiom dictionary or other idiom resource, it is considered an idiom” (p. 6). Relative idiomaticity of a particular idiom (based on internal structure), as discussed in the previous chapter, will not be considered or analyzed, but could certainly have a bearing on raters’ judgements of idiomatic versus open-choice usage.

Idiom Source

The source of idioms for the current study was the same as that used by Lynn (2016)—a compilation of 27,000 idioms from multiple English dictionaries and other electronic resources. After processing in Excel, the list had no exact duplicates, but it did contain different realizations of the same idiom (e.g., *add fuel to the argument, add fuel to the controversy, add fuel to the debate, add fuel to the fire, add fuel to the flames*), which Lynn focused on in his study.

Additionally, the raw list excludes lemmatized forms of an idiom (e.g., *pay the price* appears, but
the inflected forms *pays the price, paying the price, and paid the price* do not). However, the decision was made to utilize the raw forms on the list as they provide more than sufficient numbers of idioms for the sampling purposes of the present study. Additionally, while using the raw idiom forms found on the list may be viewed as a limitation, for the purposes of this study it is considered more important to generate random samples than find every possible context of an idiom.

**Identifying a Sample Size of Idioms**

In order to determine a sample size that would be statistically acceptable, power analysis was done with an estimated standard deviation of .2. This suggested that a sample of at least 712 idioms, with 10 contexts per idiom, would provide sufficient numbers, provided that all selections (of idioms and contexts) were random. This estimated sample size was rounded up to 1,000 for good measure. Using Excel, a stratified random sampling of 1,100 idioms (approximately 25 idioms from 27 units of 1,000 idioms each) was taken from the raw idiom list of 27,000 forms. The extra 100 idioms were a precaution in case any idiom forms had to be eliminated because they did not occur in a least 10 contexts of the selected corpus, or if they were different realizations of the same idiom (e.g., *pardon my French; excuse my French*).

**Corpus Selection**

**NOW Corpus**

The NOW corpus was selected for this study of idioms because of its size, content, and recency, and is described by Davies (2016) as follows:

The NOW corpus (News on the Web) contains 4.5 billion words of data from web-based newspapers and magazines from 2010 to the present time. More importantly, the corpus
grows by about 5-6 million words of data each day (from about 10,000 new articles), or about 150 million words each month. (http://corpus.byu.edu/now/)

Because the corpus changes daily, a certain degree of general language randomization is ensured. Comprised of web-based articles from magazines and newspapers, the NOW corpus’ content is particularly appropriate for the study of idioms as it contains authentic, current language use.

**Context Generation**

Each of the 1,000 idioms was entered into the corpus query tool on the NOW interface in the raw form listed (see Figure 1). Once entered, the corpus generated random contexts (see Figure 2).

---

**Figure 1. Idiom entry**

From these contexts, 10 samples of the raw idioms used in context, generated by the corpus concordancing software, were selected for analysis. These contexts were then “copy and pasted”
into an Excel spreadsheet to be analyzed (Figure 3). This process was consistent throughout the collection of data.

Some of the raw idiom forms were discovered to contain a pronoun that varied widely (e.g., go about my business, go about your business, go about their business, etc.). In such cases...
a wildcard (*) was inserted into the core idiom in place of the variable pronoun in order to ensure that the corpus generated sufficient contexts for analysis. An example of this process is shown below. The idiom is entered into the corpus using the wildcard (go about * business) (see Figure 4) and generates various forms of the idiom (Figure 5). Additionally, it was decided that the analysis of lemmas would add additional complexity that would be better addressed in a separate study.

Figure 4. Idiom wildcard entry

Figure 5. Corpus generation of wildcard
The variation between these forms is the pronoun. The first form is selected (Figure 6) followed by the first 10 contexts. This was the only modification made to the idioms found on the list. To ensure consistency, the researcher always selected the first form of the idiom for context selection. While this choice caused one pronoun form with its contexts to be emphasized, it was felt that the need for consistency of selection outweighed variety. If the first form of the wildcard idiom did not have 10 contexts, then all instances of the first form were selected and however many additional contexts needed were gathered from the succeeding form(s). This possibility is illustrated in Figure 7 with the core idiom “keep * nose out of it,” which only has six items in the first form. Once the idiom had been selected, the actual language contexts are visible, linked to their original sources (Figure 6).

Figure 6. Idiom contexts
Figure 7. If the first idiom form has less than 10 contexts, the second form is also collected.

**Analysis of Idioms**

Finally, each context was individually examined to determine if the idiom forms were actually being used as idioms (true idioms) in the given contexts, or simply as literal phrases with idiom forms (false idioms). This dichotomous rating was completed in an Excel spreadsheet (see Figure 9). Each idiom was listed, and numbered, next to the 10 contexts pulled from the NOW corpus. The original source for each context is hyperlinked within the document so that the rater could easily reach the original source in case the corpus context was insufficient to determine idiomaticity. The rater entered a “1” next to each context that demonstrated any level of idiomaticity (“true idiom”), and a “0” next to each context that demonstrated the open-choice principle (literal meaning—“false idiom”). The rater evaluated the idiom forms based on the context in which they were embedded.
To test the reliability of the ratings, ten percent of the contexts were analyzed by a second rater. A second Excel spreadsheet was created with 100 idioms (and their 1,000 contexts), randomly
selected from the 1,000-idiom sample. Additional columns were created for the second rater to utilize, while the original rater’s numbers were hidden. Figure 10 shows an example of the side by side rating system for the first and second raters.

Due to the somewhat subjective nature of the idiom rating process, it was decided that a phrase with any idiomaticity was rated as an idiom. To be rated as “open choice,” the phrase had to be entirely literal in the context in which it was embedded. As an example, the phrase *stem the flow* is used literally in an article from the South Coast Herald: “Despite this, nothing had been done to fix the taps, which now had a pile of rocks stacked on top of the valve in an attempt to *stem the flow of water.*” Conversely, from The Australian Financial Review: “…major parties are framing their debates in this space with one eye to trying to *stem the flow of votes to Hanson’s party….*” The first context provided by the NOW Corpus uses the words *stem the flow* as open choice (false idiom) with meaning being deduced from understanding each word. The second context uses *stem the flow* with less transparency that requires the reader to have an understanding of the phrase as a whole, or as an idiom (true idiom). Another example of a false idiom versus a true idiom can be seen with the idiom *ring a bell.* In a context from the NOW Corpus, which discusses the behavior of various trained animals, it is stated that “they now treat their owner like their butler. They *ring a bell* and the human immediately delivers sustenance.” Conversely, a figurative example is found in the following: “Of course, I did not expect to ‘know’ all top flight journalists but imagined that the name should at least *ring a bell….*” (*taken from NOW Corpus).

**Interrater Discrepancy**

While the first and second rater identified idioms identically over 90% of the time, an agreement rate considered highly reliable for this type of analysis, the results of the interrater
analysis produced some examples of the difficulty of identifying how idiom structures are functioning within authentic language. By extension, the internal structure and surface realization of an idiom could also make it difficult for language users to discern if the structure is indeed idiomatic. Table 1 contains four instances of interrater discrepancy. In each case, the raters agreed in the majority of the contexts belonging to a certain idiom form, but the exceptions were somehow identified differently. In the first example of Table 1, the idiom form *country cousin* has a use that could be identified as either open choice or idiomatic. The rater had to assume that Kagiso Rabada was either the literal cousin of Chris Morris, or that the form was being used idiomatically to describe their relationship. One rater assumed the former, the other rater assumed the latter. The ambiguity of the context is not atypical of idiom usage. A second example is seen below with the idiom form *hang fire*. The first rater identified this idiom form as a false idiom because its external structure is part of the name of an event. The idiomatic meaning of *hang fire* does not seem to be augmenting the meaning of the name in any way, hence an example of open choice. The second rater, however, marked the context as a true idiom.

A third idiom example presents an interesting dilemma for the raters. *Kick up* is the designated idiom form, but in this particular context it is found within another idiomatic form *kick up the backside*. Due to this, one rater identified the form as idiomatic. The other rater considered the form to be open choice as the context did not utilize the meaning of the form *kick up* but had a literal meaning of “kick” followed by a prepositional phrase. The final example of discrepancy found in Table 1 is of the idiom form *be a bag*. Due to the proximity of the idiom *bag of wind* it is difficult to discern if the “be a” is part of the idiomatic form or not. One rater identified *be a bag* as idiomatic, while the other identified it as open choice. It is also likely that many of the
forms with interrater discrepancy are archaic and, consequently, unfamiliar to the raters, making them difficult to identify.

Table 1. Interrater Discrepancy Examples

<table>
<thead>
<tr>
<th>Idiom No.</th>
<th>Context</th>
</tr>
</thead>
<tbody>
<tr>
<td>152</td>
<td>From Australia's Pat Cummins, South Africa's Kagiso Rabada and his <strong>country cousin</strong> Chris Morris to India's Mohd Shami…</td>
</tr>
<tr>
<td>298</td>
<td>…light bunches of joss sticks to place into the' dragon' during the Tai <strong>Hang Fire</strong> Dragon Dance in Hong Kong on Sept. 14, 2016.</td>
</tr>
<tr>
<td>446</td>
<td>…even if Lee Cattermole is back to give the dressing room an extra <strong>kick up</strong> the backside.</td>
</tr>
<tr>
<td>649</td>
<td>I suspect that Mmusi has turned out to <strong>be a bag</strong> of wind. He is hardly heard and that reminds me of Mbeki's silent…</td>
</tr>
</tbody>
</table>

Interestingly, such differences in judgement point out the fuzzy line that often exists between literalness (open choice) and nonliteralness (idioms) because of the contexts in which phrases are embedded. If this kind of incongruity is happening with native raters, what does it say for the problems facing language learners?

Despite the difficulty in the language chunks as idioms, or false idioms, the disagreement was acceptably low and supports the validity of the data analysis. All ratings were done independently. When any disagreement arose, the raters discussed the reasons for their decisions, but it was determined in the end to maintain the original ratings from first rater (researcher).
Chapter 4: Results and Discussion

The stated purpose of this study is to discover to what degree language chunks, with idiom surface structures, may be relied on to actually act as idioms in authentic contexts. Altogether, 1,000 core idioms were entered into the NOW corpus to collect 10 authentic language contexts per idiom. As discussed in the methodology section, these 10,000 potential-idiom contexts were then analyzed for idiomaticity, each language sample being marked based on a dichotomous judgement (i.e., 1 = idiom; 0 = open choice)

Results

The results of this analysis showed that over 79% of the idioms sampled (see Appendix A) were functioning idiomatically in all ten of their randomly selected contexts, with nearly 92% of the idioms functioning idiomatically in at least half (5+ contexts) of their randomly selected contexts. These numbers are demonstrated in Figure 11, which shows what percentage of the analyzed idioms functioned idiomatically in 10 out of the 10 contexts examined, as compared to the percentage of idioms that functioned idiomatically in 9 out of the 10 contexts examined, 8 out of 10, 7 out of 10, and so forth.

An important variable to take into consideration with these numbers is the source from which the contexts were drawn. As noted in Chapter 3, the NOW corpus was selected for its size, recency, and authentic language, all ideal for the purposes of this study. The corpus does not, however, include extensive spoken language, a register that utilizes idiomatic language extensively. The corpus also draws upon international sources so that the contexts of this study are not necessarily characteristic of a single form of regional English. Such variables could possibly change the results to some degree, but there is also strong reason to believe that the basic findings are reliable. In short, the data shows that the analyzed idiom forms function
overwhelmingly as true idioms in context. The data also reveals interesting trends across the contexts. The following sections will discuss details relating to idiom forms functioning as idioms, idiom forms functioning as open-choice language, and idiom forms functioning as proper nouns.

Figure 10. Degree of Idiomaticity (Note: rounding caused the total percentage to be less than 100)
Idiom Principle (True Idioms)

The overwhelming trend in the results of this study revealed language chunks identified as idioms being used idiomatically in actual usage. As anticipated, these language chunks represented a single choice on the part of the writer to convey a figurative message. Examples from the NOW corpus include “But let's not lose sight of the wood for the trees” and “…not sure how much I'll be able to write today before I flake out.” Both of these examples show idioms being used in a way that requires understanding of the phrases themselves. Additional examples, all of which were found to function idiomatically in 100% of the sampled contexts (10 out of 10 per idiom), can be found in Table 2. These have been referred to as “true idioms” when they function idiomatically in actual usage.

Table 2. Idiomatic language examples.

<table>
<thead>
<tr>
<th>Idiom No.</th>
<th>Idiom</th>
<th>Context</th>
</tr>
</thead>
<tbody>
<tr>
<td>148</td>
<td>let it slide</td>
<td>him to do that, I was even willing to let it slide after I saw</td>
</tr>
<tr>
<td>268</td>
<td>lift the spirits</td>
<td>hint of bonfire smoke in the air all combine to lift the spirits. Senses are sharpened too by the knowledge that the</td>
</tr>
<tr>
<td>297</td>
<td>high horse</td>
<td>One told me to get off my highhorse. Another said she was unsure why I was complaining given nothing even happened</td>
</tr>
<tr>
<td>369</td>
<td>heart sinks</td>
<td>I remember, even as it starts to fade. My heart sinks, as it did for the past so many months. I came to America</td>
</tr>
<tr>
<td>420</td>
<td>pick-me-up</td>
<td>this shaved ice dessert is a perfect pick-me-up for those seeking respi the heat</td>
</tr>
<tr>
<td>590</td>
<td>throw the book at</td>
<td>personnel who dishonor their uniforms and stain their badges. We will not hesitate to throw the book at such hardheads.</td>
</tr>
<tr>
<td>629</td>
<td>all hell break loose</td>
<td>concerned about the morale of his co-workers. &quot; We have seen all hell break loose, &quot; he says. &quot; Families and patients do not</td>
</tr>
<tr>
<td>793</td>
<td>whitewash</td>
<td>The myriad of institutional forces that seek to whitewash ethnic mino stories must be dismantled, especially considering the</td>
</tr>
<tr>
<td>910</td>
<td>under the pump</td>
<td>It wasn't always in our favour. The Chiefs had us under the pump for many parts of the game, but the things that we had</td>
</tr>
<tr>
<td>927</td>
<td>rat race</td>
<td>to find a way out of the rat race then these eccentric cafes are just the thing for you. Book cafes today</td>
</tr>
</tbody>
</table>
**Open-Choice Principle (False Idioms)**

The analysis also revealed that only 23 of the 1,000 idioms sampled (2.3%) had no idiomatic representation in the 10 contexts that were sampled for each. Eight of these 23 language chunks are found in Table 3, along with a sample context of the idiom form being used in an open choice way; a complete list of these 23 idiom forms can be found in Appendix B. These have been referred to as “false idioms,” when they do not function idiomatically in actual usage. This is not to say that they never function idiomatically, only that they did not in the contexts analyzed. Some interesting factors may have affected the particular 23 that never functioned as true idioms in their 10 analyzed contexts. One possible factor is that the genre of writing the NOW corpus predominantly represents—news articles—tends to focus on factual information. Consequently, some idioms occur as open-choice in this genre that would more likely be idiomatic in another register of language. For example, *throw acid in her face* is literal in all 10 contexts pulled from the corpus, with all 10 of those contexts dealing with physical attacks utilizing acid.

Another interesting commonality of false idioms in the analysis is the number of compound nouns. Many of the idiom forms that were also compound nouns had a lower frequency of idiomatic usage. Examples of these compound nouns include: *diamond ring* (“she is wearing a massive *diamond ring*”), *outer core* (“limits of what some geophysicists think Earth's *outer core* is capable of doing”), and *chemical imbalance* (“an illness in the body or diabetes is a *chemical imbalance* in the body”).
<table>
<thead>
<tr>
<th>Idiom No.</th>
<th>Idiom</th>
<th>Context</th>
</tr>
</thead>
<tbody>
<tr>
<td>529</td>
<td>elephant ear</td>
<td>… up with a guncase and a handbag made from elephant skin. There is an elephant ear in the farmhouse, and an elephant foot cover by zebra skin.</td>
</tr>
<tr>
<td>649</td>
<td>be a bag</td>
<td>afternoon after authorities discovered a suspicious package that was later determined to be a bag of children's toys.</td>
</tr>
<tr>
<td>766</td>
<td>elder brother</td>
<td>… decapitated herself in front of Goddess Kali's idol, SP Biswas added. The elder brother rushed to see his mother with Narayan and saw his mother's severed head</td>
</tr>
<tr>
<td>788</td>
<td>brick in</td>
<td>…family that will make you howl louder than when you accidentally tread on a Lego brick in the dead of the night.</td>
</tr>
<tr>
<td>879</td>
<td>on the house</td>
<td>A Fannie Mae spokesman said there was no insurance on the house, which led to an agreement to sell it for $50,000.</td>
</tr>
<tr>
<td>931</td>
<td>throw acid in her face</td>
<td>After her ex-boyfriend Daniel Lynch hired Stefan Sylvestre to throw acid in her face in a brutal attack, she has been forced to undergo</td>
</tr>
<tr>
<td>959</td>
<td>sour cherry</td>
<td>There are tomatoes with purple basil olive oil, hung cheese with sour cherry compote, wood-roasted tahini with caroub molasses, candied pumpkin yoghurt, grilled halloumi</td>
</tr>
<tr>
<td>966</td>
<td>cheese off</td>
<td>…preparing for an exam, then the next thing you know, you're licking cheese off your fingers or shaking off cookie crumbles on your lap.</td>
</tr>
</tbody>
</table>

A final characteristic of these “low-performing” idioms is brevity. Many of the idioms made up of a single word, or a single word and a common lexical item such as a preposition or a determiner, had fewer idiomatic occurrences. Examples of this are some people, to a man, cheese off, and rule in. The more common forms of these idioms made them more likely to be used in an open-choice manner (literally), even though they can be used in idiomatic contexts. Nevertheless, they are less reliable as idioms due to their more changeable nature (*examples taken from NOW Corpus).
Idioms as Proper Nouns

Finally, as the analysis progressed, it was interesting to observe the frequency with which idioms functioned as proper nouns (see Table 4). The researcher decided that these proper nouns are indeed functioning idiomatically because they are being utilized as proper nouns as a result of their idiomatic meaning. Thus, their idiomaticity is inherent to their usage because the connotation of meaning was deliberate and significant.

Table 4. Idioms functioning as Proper Nouns

<table>
<thead>
<tr>
<th>Idiom No.</th>
<th>Idiom</th>
<th>Context</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>golden hello</td>
<td>New applicants can also be awarded a 'Golden Hello' of up to 10,000, whilst the new Junior Entry Apprenticeship for young…</td>
</tr>
<tr>
<td>28</td>
<td>sleepyhead</td>
<td>Travelling Bugs, Rapping Beavers, and funny characters like Sir Bucket Head, Princess Sleepy Head, speedy the Ant…</td>
</tr>
<tr>
<td>63</td>
<td>step by step</td>
<td>Read more Sandra Nelson, manager of Step by Step Nursery in Bridge Avenue, Hammersmith, where road emissions levels are</td>
</tr>
<tr>
<td>84</td>
<td>throw me to the dogs</td>
<td>The 21-year-old's latest project comes after his previous film, Throw Me to the Dogs, won five awards and was screened at</td>
</tr>
<tr>
<td>89</td>
<td>happy hour</td>
<td>Stand Up for the Week, Al Murray's Happy Hour, All Star Mr and Mrs, Celebrity Who Wants To Be A Millionaire?</td>
</tr>
<tr>
<td>91</td>
<td>rolling stone</td>
<td>March, Keith Richards paid tribute to his musical idol for the upcoming issue of Rolling Stone. Chuck Berry once gave me a</td>
</tr>
<tr>
<td>132</td>
<td>thick as a brick</td>
<td>Albums like Aqualung and Thick as a Brick cemented Jethro Tull's status as one of the most popular bands of the 1970s</td>
</tr>
<tr>
<td>166</td>
<td>saved by the bell</td>
<td>…be fronted by none other than Mario Lopez. Our favourite Saved by the Bell actor will be perfect for the presenting role.</td>
</tr>
<tr>
<td>191</td>
<td>signed sealed delivered</td>
<td>…later however, channeling the Motown sound with their version of Stevie Wonder's Signed Sealed Delivered that</td>
</tr>
<tr>
<td>192</td>
<td>living end</td>
<td>Chris Cheney of the Living End plays St Jimmy, a role which will be taken over by Grinspoon's</td>
</tr>
</tbody>
</table>

Examples of this are the idiom rolling stone used as the name of a magazine, The Rolling Stone, as well as the idiom funny bone which is seen as the stage name of a performing comedian.

Multiple idioms (throw * to the dogs, prick up your ears, and saved by the bell) are seen as the
title of films or television shows (“Throw Me to the Dogs”, “Prick Up Your Ears”, and “Saved By the Bell”). Idioms often presented themselves as song titles, such as the song “Hit Me With Your Best Shot” by Pat Benatar. This pattern of idioms being used as proper nouns (i.e., song titles, pseudonyms, movie titles, etc.) shows the significance of idiomatic language in popular culture and the necessity of comprehending idioms for those who wish to fully grasp the pragmatics of the language.

Discussion

The findings of this study suggest an overwhelming tendency for structures identified as idioms to, in fact, act as idioms in actual contexts, with nearly 80% of the idioms studied functioning as idioms in all contexts analyzed, and 94% of the total idioms studied functioning idiomatically at least half of the time. Consequently, it may be assumed that these language chunks, identified as idioms, will most likely function as such in the language.

The ramifications of this are far reaching. First, while some idioms have literal counterparts in actual language usage, the vast majority do function as idioms and can therefore be taught and researched as lexical items, rather than strings of words that may or may not be idiomatic in actual usage. Second, in one sense, this study is another validation of frequency-based corpus research that has found idiomatic strings and other multiword items to be prevalent in the language. No major study until now has actually tested the form-based data to determine how well it matches context-based meaning. There have been studies focusing on variations in idiom forms (e.g., Geyken, 2015; Moon, 2015), but none, to the knowledge of the researcher, dealing with the form-versus-meaning analysis performed in this study. Third, given the fact that most idiom structures do function idiomatically in context, there is even more reason not to process language on a word-by-word basis as has been done for many years based on grammar-
heavy approaches to linguistic computing, or to teach only individual words in language training. Finally, the fact that some idiom structures do not function idiomatically very often in actual usage suggests that a rating system might be in order to assist both language teaching and linguistic computing. In other words, idiom structures that are both frequent and function primarily as idioms should be taught before those that are less frequent and/or that appear less often as actual idioms in context. Such a rating system also has ramifications for linguistic computing in terms of the confidence that programmers should have in marking or tagging idioms as idioms.
Chapter 5: Conclusion

The findings of this study, which measures the frequency with which supposed idioms act idiomatically in the language, reiterate the significance of idioms in language pragmatics and support the view that idioms are “semi-preconstructed phrases that constitute single choices” (Sinclair, 1991, p. 110, emphasis added). With over 90% of the idiom forms studied functioning idiomatically the majority of the time, it is evident that these “preconstructed phrases” are indeed representative of a single choice for the language users. Consequently, comprehension of these language units is important for a robust and native-like vocabulary. This finding has important ramifications for the acquisitions of idioms, as learners can have a high degree of confidence that idiom forms they encounter and spend time learning are rarely found in literal usage. If this were a 30-70 (idiom vs. open choice) proposition, or even a 50-50 proposition, rather than a 90-10 proposition, it would make the learning task much more difficult, and perhaps even hard to justify.

Likewise, corpus and other researchers who attempt to account for idioms in electronic text can and should begin to link idiom forms together as phrasal vocabulary items, rather than processing them as individual grammatical words, as has been done in the vast majority of cases to date. One way to look at this is to ask if we would be happier to be accurate 10% of the time or 90% of the time in identifying, counting, and reporting idiom data.

However, this study also shows that while most idiom forms function idiomatically nearly all the time, others rarely do, thus confirming the importance of considering the frequency with which specific idiom forms function idiomatically. Further research like that performed in this study has the potential of establishing idiom forms on a scale of relative “idiomaticity”—that
is, those with the highest and lowest possibilities of acting idiomatically, and everywhere in between. The value of such a scale for teaching, learning, and research is clear.

**Limitations**

Limitations to this study that are important to note are as follows: First, the idiom sample consisted of the raw form of the idioms rather than their lemmatized form. This means that variants (e.g. *losing his head, will lose his head*) of the idiom’s form (*lost his head*) as identified in the list were excluded. Another example of this situation is seen with the idiom *shake the foundation*, the form used in this study, which excludes the past tense form *shook the foundation*. The data analysis utilized the idiom forms exactly as they were found on the list, or in their raw form. However, there is no reason to believe that the substantive findings of this study would be altered if such a lemmatized approach had been taken.

Second, the overwhelming magnitude of English idioms necessitated that only a practical sample size be utilized in the present study, and there is always the chance that results would have been different if more idioms had been analyzed. However, it is felt that this limitation was largely accounted for by randomly sampling a statistically viable number of idioms from the most comprehensive list available.

A third limitation is that the idiom list utilized in this study consists only of idioms found in more traditional idiom dictionaries and other resources, thus excluding some idioms of a more colloquial nature—for example, those characteristically present in “street talk” and very informal settings, many of which have vulgar connotations. Some of these idioms are likely to be found in resources like Urban Dictionary (http://www.urbandictionary.com/), but the decision was made not to use such resources because of the intended audience of this study, the nature of the NOW corpus used in the analysis, and the difficulty of finding a large sample of colloquial
spoken English where many of these informal idioms would be found. A related limitation is that new idioms and idiom variations are always entering the language as the language grows and evolves. These newest forms could appear in the NOW Corpus which is updated daily, but they will not be found on the list of canonized idioms from already published sources.

Finally, the corpus itself was limited to online newspapers and magazines, excluding spoken contexts as well as wider genres of language. The NOW corpus was selected as the most appropriate corpus for the present study for its immense volume of language as well as its perpetually current status, and while the corpus and idiom list exclude potentially valuable spoken language samples, it was a limitation that had to be accepted because of the limited availability of authentic spoken samples. Despite these limitations, it is felt that this study has produced important data that may be utilized to augment teaching, learning, and further research.

Extensions

As the first of its kind, this study lends itself to many new applications and extensions for linguists and teachers alike. Idioms have long been an acknowledged hurdle for ELLs due to their phrasal nature and non-literal meanings. They are at once “notoriously difficult” for language learners and also “a great asset to learner[s] in acquiring a new language” (Celce-Murcia & Larsen-Freeman, 1999, p. 39). Consequently, this study can do much to inform the development of instructional content and language teaching by:

1. Raising professional awareness of the consistency with which idioms are functioning idiomatically and supporting this with quantifiable data, thus encouraging the direct instruction of idioms.

2. Informing instructors and developers regarding the consistent nature of the contexts in which idioms are embedded, thus validating the single-choice nature of English idioms.
3. Highlighting those idioms from the random sample of the study that were highly idiomatic, appearing in all contexts as actual idioms. These could be used for a foundation for selecting idioms for direct instruction.

Additionally, this data provides a new approach to the linguistic analysis of idioms, by actually testing the claim that idioms do predominantly represent a single choice by users of the language and not an open choice. Corpus and computational linguists can use the data in this study as justification for the tagging of idiomatic forms in linguistic computing.

**Future Research**

Possibilities for further research include: (1) a replication of this study done with different registers of the language (e.g., spoken) to compare with the results of the current study dealing with online newspapers and magazines. Emphasis could also be placed on idioms found in English for specific purposes scenarios such as business, biology, athletics, and education; (2) a replication of this study using more inclusive search criteria of idioms in their contexts (lemmatized forms, variant forms, etc.)—i.e., not restricting the search to core forms—to extend and clarify the findings of the current study; (3) a more nuanced investigation of idioms functioning as proper nouns—one of the unexpected findings in the current study.

Finally, these are very complex undertakings and the present study has been just one step in measuring the reliability of idiomatic forms to act reliably as idioms in the English language. Further investigation is warranted to shed additional light on this important and timely avenue of study.
References


Appendix A

Idioms with High Idiomaticity (10/10)
two in the bush
deal * in
roll **

blue-sky thinking
miss the point
wing it

take umbrage
hatchet man
life and soul (of the party)

golden hello
back*seat driver
second wind

prime the pump
turf war
go to blazes

high as a kite
bone war
easy on the ear

panties in a bunch
go about * business
would give * eye

stretch the truth
dead loss
prick up * ears


going rate
find * voice
onto a good thing

tight spot
hidden agenda
make do and mend

tilt at windmills
Buckley's chance
break/bust * balls

drop * drawers
rabbit hole
toss around

flash in the pan
broad-brush
do the deed

brown thumb
keep * nose out of sth
cut and thrust

good run
long in the tooth
at pains

match day
step by step
put * foot down

take the field
come out of left field
keep tabs

final/last straw
make hay
turn * hand to

common law
hog heaven
take * life into *

like wildfire
buck the system
know where * stands

twist * words
hit list
wolf in sheeps' clothing

off guard
ring true
put/set the record straight

do * homework
(right) on the button
pot calling the kettle black

sleepy head
with * own fair hands
go wild

crack up
fill the gap
a run for * money

piece de resistance
false step
worlds apart

sheets to the wind
funny bone
libel chill

town and gown
heck of a *
my word

worm food
hit list
mover and shaker

pick * pocket
ring true
how's it hanging

give * a * hand
(right) on the button
Bible-basher

watch * language
false step
thick as a brick

go downhill
funny bone
go out of * mind

soldier of fortune
heck of a *
no stomach for *

go to/take great pains
follow in * footsteps
generation gap

catch * rays
do * part
happy hour

busman's holiday
don't let the bedbugs bite
in these parts

roll **
six ways to Sunday
meals on wheels
battle of wits
hard line*
firm hand
let * slide
make or break
country cousin
corridors of power
rotten in the state of Denmark
pack a punch
on a tear
take the money and run
come to * aid
walk in on
straight cash
get/start off on the * foot
saved by the bell
fired up
have a new lease on life
rock/shake the foundation
broaded/widen/expand *
horizons
chill down * spine
wood for the trees
new blood
good day
brass-necked
soul mate
goosed
flake out
shut * eyes to
set of wheels
find * way
break * duck
now * mention it
take the liberty
spoiled for choice
signed, sealed, delivered
living end
* wish is * command
wonder of wonders
talking head
alive and kicking
one-horse town
creature of habit
out of action
fall short
bag of wind
shit-stirrer
spur of the moment
backpedal
money is no object
head/run/take for the hills
fall into a trap
head start
from the bottom of * heart
suit *
off the top of * head
sound/dead/fast asleep
run * course
grounded
pill mill
nerves of steel
get/give * carte blanche
play the field
compare notes
glamour girl/puss
keep the edge
show */have a good time
shack up
skinny dip
run for office
break out in a rash
bring * to light
track record
a feel for *
no two ways
like I said
kill * with kindness
end of the rainbow
no contest
fact is
drop charges
against the current
strike * fancy
fall into * hands
in years
tied to * apron strings
business as usual
box * ears
punch-drunk
tough as old boots
amateur hour
chance your arm
screw the pooch
more's the pity
two-second rule
dead hand
pick * up
raise/lift * spirits
extra pair of hands
on good/bad terms
good question
too big for * britches
talent management
pour/throw cold water on
swing a cat
take it like a man
break * back
eat like a horse
mud in * eye
cry * eyes out
mission creep
sucker punch
bitter pill
carry the day
accidents will happen
back of beyond
double booked
beat a retreat
keep * eye in
high horse
eat * heart out
blue devil
showdown
millstone around * neck
world at * feet
fall apart
head trip
good old boy
itsy-bitsy
call the tune
flying colo(u)rs
catch a train
turn a blind eye
fall flat
dip out
put to the test
no match for *
sweet tooth
beat the system
gut level
frog in * throat
step on the gas
call off the dogs
door to door
a life of its own

not feeling *
pin down *
move up/down in the world
serious coin
blacken * name
miss *
shady deal
good drunk
jump * bones
judgment call
quantum jump
go/move up the ladder
cold/hard cash
hold * head high or head held
high
rumor mill
on tenterhooks
in a league of * own
all in a day's work
shouting match
throw down the gauntlet
light on * feet
pick-and-mix
cool cat
guard * tongue
care/give a toss
future looks bright
lead the line
cash cow
settle a score
turn a deaf ear
light of day
heart sinks
on edge
go down/up in flames
fall from grace
public enemy number one

breathe easy/ier
low-hanging fruit
back to the grind
small talk
fool's errand
pick up the pieces
lick * wounds
back-breaking
Dunkirk spirit
rug up
check * out
loved up
throw * weight around
loose change
faint of heart
do a number on *
slave to fashion
turn to jelly
cut to the chase
tempt fate
put flesh on * * *
fish bowl
if/when it comes to the crunch
error of * ways
spear carrier
lotus land
stop the presses
iron fist in a velvet glove
ring off the hook
* guess is as good as mine
kick the bucket
man of the cloth
pick-me-up
muddle along
good as gold
play with fire
shed tears
hang up * boots
middleman
argue the toss
side of the coin
take * hate off to
good old days
good morning
throw down * arms
stars are aligned
pay through the nose
fools rush in
tangled web
safe ground
grass widow(er)
trip the light fantastic
steam coming out of * ears
shoot * in the foot
ready cash/ money
tighten the purse strings
all bark and no bite
on par
notch above/below
cup of joe
all good things come to an end
snowball effect
not take no for an answer
catch/cop/get some z's
* whole world *
from/out of the top drawer
ask/cry for the moon
guilt trip
bottom of the heap/pile
move with the times
put the heat on *
fruit of the poisonous tree
to a nicety
kiss up
stock phrase
give the devil * due
mop/wipe the floor with *
hit the fan
if memory serves correctly
put to the sword
acquired taste
push *
cool customer
change tack
above the law
actions speak louder than words
life is cheap
johnny-come-lately
hole*in*the*wall
face of the earth
tie * in knots
choke a horse
dog in the manger
black-listed
blue-collar
be all things to all men
nest egg
drugstore cowboy
forbidden fruit
play hookey/hooky
quarter-pounder
hanging in the air
copy cat
a kick in the teeth
layman's terms
beat/play * at * own game
time flies
pick of the litter
be still my heart
beat * to a pulp
right on the money
bring * to * knees
kiss and cry
such is life
low-end
second thoughts
end justifies the means
stay tuned
bone idle
have/take a seat
honest to God
game is up
hard pill
beast of burden
beyond/past the end
grab a chair
street appeal
head and shoulders above
luck may/would have it
settle * stomach
in * line
call * into question
speed merchant
throw the book at
middle-aged spread
out of/from the corner of * eye
passing fancy
shut * mouth
middle of nowhere
traditional marriage
clutch/grasp at straws
kick in the balls
welcome to the club
paper trail
in that vein
make the world go round
read/see the handwriting
fickle finger of fate
lord of the flies
hard on the eyes
every man Jack *
installed base
by virtue of
have a ball
give * a break
man of the hour
living hell
keep pace
break * word
pump up the volume
busy beaver
sweat buckets
long finger
barrow man
article of faith
all hell break loose
cabin fever
take a picture
not even funny
get a job
let it all hang out
look upon
win the day
big break
pull the plug
olde-wordle
* like hell
poetic license
as the next guy
worship the ground *
old flame
wash * dirty
pearls before swine
fat*cat
feast or famine
risky business
hand it to
happy-go-lucky
gnash* * teeth
clip * wings
day and age
bodice-ripper
whoop-ass
straw that stirs the drink
bad hair day
man of * parts
do * by halves
work up a thirst
suck ass
dirty look
have fun
ground-breaking
strong, silent type
turn a hair
best-kept secret
down in the mouth
trade secret
snowed under
go potty
white as a ghost
Joe Citizen
foregone conclusion
horse opera
stand behind
can do * in * sleep
heavens to Betsy
break bread
two nickels to rub
none the wiser
white coat hypertension
stars in * eyes
<table>
<thead>
<tr>
<th>Honor-bound</th>
<th>Wait on hand foot (and finger)</th>
<th>Whitewash</th>
</tr>
</thead>
<tbody>
<tr>
<td>Done deal</td>
<td>Horror show</td>
<td>Culpable homicide</td>
</tr>
<tr>
<td>Come to light</td>
<td>Non-denial denial</td>
<td>Tie the knot</td>
</tr>
<tr>
<td>Time to time</td>
<td>Take a back seat</td>
<td>Return compliment</td>
</tr>
<tr>
<td>Tin ear</td>
<td>Cling-on</td>
<td>Sinking feeling</td>
</tr>
<tr>
<td>Lost for words</td>
<td>Death spiral</td>
<td>Eye of the beholder</td>
</tr>
<tr>
<td>Trailer (park) trash</td>
<td>Own man</td>
<td>Work up</td>
</tr>
<tr>
<td>Out of the spotlight</td>
<td>Cutting edge</td>
<td>Boys will be boys</td>
</tr>
<tr>
<td>Tickle * fancy</td>
<td>For the time being</td>
<td>Rub salt in wound</td>
</tr>
<tr>
<td>Grass is (always) greener</td>
<td>A moment's notice</td>
<td>Growing pains</td>
</tr>
<tr>
<td>Put the hammer down</td>
<td>Boys toys</td>
<td>Have the heart to</td>
</tr>
<tr>
<td>Off * pedestal</td>
<td>Balls to the wall</td>
<td>Take the point</td>
</tr>
<tr>
<td>Do duty as *</td>
<td>With a jaundiced eye</td>
<td>Run hot and cold</td>
</tr>
<tr>
<td>Hit * with * best shot</td>
<td>* the screws on</td>
<td>Scales fall from eyes</td>
</tr>
<tr>
<td>Borrowed time</td>
<td>Free as a bird</td>
<td>Separate/sift/sort the wheat</td>
</tr>
<tr>
<td>In the palm of *</td>
<td>Faceless bureaucrat</td>
<td>Beauty queen</td>
</tr>
<tr>
<td>In the works</td>
<td>Strain every nerve</td>
<td>Stick out</td>
</tr>
<tr>
<td>Ride shotgun</td>
<td>* thought as much</td>
<td>With sugar on top</td>
</tr>
<tr>
<td>Leopard can't change its spots</td>
<td>Lounge lizard</td>
<td>Blot copybook</td>
</tr>
<tr>
<td>Bad time</td>
<td>Shipshape</td>
<td>Purple prose</td>
</tr>
<tr>
<td>Hold up * end</td>
<td>Grey matter</td>
<td>In light of</td>
</tr>
<tr>
<td>Lightning-quick</td>
<td>So help me god</td>
<td>Pre-menstrual syndrome</td>
</tr>
<tr>
<td>Good job</td>
<td>Checkup</td>
<td>Hundred and ten percent</td>
</tr>
<tr>
<td>Blood is up</td>
<td>God/heaven help</td>
<td>Kick in the stomach</td>
</tr>
<tr>
<td>Won't tell a soul</td>
<td>Shift * ground</td>
<td>Stick nose in</td>
</tr>
<tr>
<td>Catch a bus</td>
<td>Eat * shorts</td>
<td>Dates</td>
</tr>
<tr>
<td>Receiving end</td>
<td>Remain nameless</td>
<td>Talk out</td>
</tr>
<tr>
<td>Just a titch</td>
<td>Miner's canary</td>
<td>Take a joke</td>
</tr>
<tr>
<td>Snap to it</td>
<td>Paid-up member</td>
<td>Bring down the house/bring</td>
</tr>
<tr>
<td>Take the mick(ey)</td>
<td>Lock * up</td>
<td>Down the house</td>
</tr>
<tr>
<td>Not say a word</td>
<td>In virtue of</td>
<td>At the bit</td>
</tr>
<tr>
<td>Take * word for it</td>
<td>Dry run</td>
<td>At glance</td>
</tr>
<tr>
<td>Which end is up</td>
<td>Pack * in</td>
<td>Hell<em>for</em>leather</td>
</tr>
<tr>
<td>Hen night/party</td>
<td>Right away</td>
<td>* score and * years</td>
</tr>
<tr>
<td>Look * up</td>
<td>Checks and balances</td>
<td>On * bended knee</td>
</tr>
<tr>
<td>Carry * weight</td>
<td>Scot-free</td>
<td>Keep * bended knee</td>
</tr>
<tr>
<td>Laundered money</td>
<td></td>
<td>Take part</td>
</tr>
</tbody>
</table>
bang to rights
dirt nap
in agreement
ride/roll with the punches
nut out
flight of fancy
story of my life
snake out
pass the time of day
put the boot in
sticking point
rat race
legal duty
take to * heels
window of opportunity
shooting fish in a barrel
* not before time
worth the paper
jack of all trades
final cut
at first blush
under * nose
Elysian fields
third hand
cold-turkey
sleep with the fishes
* face is a picture
cry/sob * heart out
too rich for * blood
piece of ass
throw a game
trial run
real time
best/worst of both worlds
sign in/off/on/out/up
daylight robbery
green around/about the gills
hang * head
with an eye towards
hang * head
dice are loaded
a man about-town
first/second fiddle
rocket science
fall guy
go bust
miss the mark
death blow
blaze a trail
finger on the pulse
run roughshod
course of time
time is ripe
man of the moment
show on the road
whale of a *
not long for this world
an end in itself
hopes dashed
no time for *
wheels in motion
under the pump
play the stock market
bottom line
fly* blind
at the hands of
burn * boats
out of a sow's ear
whip hand
answer to * prayers
burn * boats
control/rule with an iron fist
whip hand
Bible-thumper
breath of fresh air
poisoned chalice
all that glitters is not gold
clever clogs
take a crap
cut corners
take a crap
take to * heels
take to * heels
* not before time
<table>
<thead>
<tr>
<th>golf widow</th>
<th>common cause</th>
<th>jug ears</th>
</tr>
</thead>
<tbody>
<tr>
<td>brick by brick</td>
<td>happy accident</td>
<td>enter/get into the spirit</td>
</tr>
<tr>
<td>two-way street</td>
<td>fall to bits</td>
<td>do drugs</td>
</tr>
<tr>
<td>start out</td>
<td>not * for the world</td>
<td>lighten up</td>
</tr>
<tr>
<td>yeoman's service</td>
<td>dead to the world</td>
<td>stick to beat * with</td>
</tr>
<tr>
<td>set sail</td>
<td>go to pot</td>
<td>spin * wheels</td>
</tr>
<tr>
<td>I just/just about/almost</td>
<td>seal * fate</td>
<td>at * expense</td>
</tr>
<tr>
<td>died</td>
<td>bits and pieces</td>
<td>for the ages</td>
</tr>
<tr>
<td>turning point</td>
<td>farmer's tan</td>
<td></td>
</tr>
</tbody>
</table>
## Appendix B

### Idioms with Low Idiomaticity (0/10)

<table>
<thead>
<tr>
<th>Idiom</th>
<th>Idiom</th>
<th>Idiom</th>
</tr>
</thead>
<tbody>
<tr>
<td>take a number</td>
<td>rule in</td>
<td>brick in</td>
</tr>
<tr>
<td>some people</td>
<td>elephant ear</td>
<td>on the house</td>
</tr>
<tr>
<td>have a pair</td>
<td>outer core</td>
<td>throw * in * face</td>
</tr>
<tr>
<td>ground sloth</td>
<td>chemical imbalance</td>
<td>sour cherry</td>
</tr>
<tr>
<td>miles away</td>
<td>diamond ring</td>
<td>cheese off</td>
</tr>
<tr>
<td>black man</td>
<td>be * bag</td>
<td>* baby</td>
</tr>
<tr>
<td>wear * hat*</td>
<td>show * butt</td>
<td>buttonhole *</td>
</tr>
<tr>
<td>to a man</td>
<td>elder brother</td>
<td></td>
</tr>
</tbody>
</table>