ITL
International Journal of Applied Linguistics

HOW MANY IDIOMS ARE THERE IN ENGLISH?

Lynn Grant and Paul Nation

ABSTRACT

The word idiom and its derivatives idiomatic, idiomatically and idiomaticity are used with a wide range of meanings. Idiomatic English is used to refer to fluent language use that sounds like that of a native speaker. Items loosely classed as idioms include colloquial expressions, collocations, acceptable but unusual expressions, and opaque multi-word units. If linguists are pressed to define what an idiom is, they usually say that an idiom is a multi-word unit where the meaning of the whole unit is not clear from the meaning of its parts. The purpose of this paper is to report on a study where one particular use of the term idiom was very carefully defined and to show what the effects of the application of this careful definition were on coming up with a definitive list of idioms. It is argued that carefully distinguishing idioms from other multi-word units makes sense for the teaching and learning of multi-word units because different approaches are needed for the different types of multi-word units. Phrasal verbs were not included in the study.

Core idioms

The two criteria used in the study to decide if a multi-word unit (MWU) was a "core idiom" were compositionality and figurativeness (Grant and Bauer, 2004). For example, *by and large* is a core idiom because the meaning of *by and large* cannot be related to the meaning of the individual words *by* and *large*, that is, it is non-compositional, and we cannot visualise some figurative use of *by and large* relating it to its meaning. On the other hand, *the worm turns* is not a core idiom. We can imagine or visualise a mild worm turning fiercely on an attacker (which conveys the meaning of the phrase) and so it is figurative and thus not a core idiom. The term *core idiom* was used to distinguish it from the looser uses of *idiom*. The test for compositionality was whether the meaning of the multi-word unit could be gained from the meaning of its parts. Just using this test alone would mean that units like *put down roots, be in the dark*, and *a hidden agenda* would be classed as core idioms, although it is clear that with some thought a person could work out their meaning with the help of context and using general cognitive principles. Thus a second test was needed to distinguish these figuratives from core idioms. The test for figurativeness was to see if the unit could be pragmatically reinterpreted to make sense in the sentence.

The procedure was as follows. It is explained in greater detail in Grant and Bauer (2004). The questions are asked in this order.

1. Is the meaning of the MWU retained if you replace each lexical word in the MWU with its own definition?

   Yes = compositional a literal 'sequence   No = non-compositional

2. Is it possible to understand the meaning of the MWU by recognising the untruth and pragmatically reinterpreting it in a way that correctly explains the MWU?

   Yes = a figurative   No = non-figurative

3. Is there only one word in the MWU which is either not literal or non-compositional?

   Yes = a ONCE   No = a core idiom

If we answer NO to these three questions, the MWU is a 'core idiom'.
If we answer YES to any of these three questions, the MWU is *not* a 'core idiom'. It may be a literal sequence, a figurative, or a ONCE. There will be some borderline cases. A ONCE is a multi-word item where one of the words does not have its normal meaning. It is called a ONCE because it is like a "once-off" use of the word.

Here is the test applied to some examples.

Head over heels in love
1   Is it compositional? NO
2   Are we able to recognise the untruth and pragmatically reinterpret it as a truth? YES
Therefore, it is NOT a core idiom, it is a figurative.

the dog days
1   Is it compositional? NO
2   Are we able to recognise the untruth and pragmatically reinterpret it as a truth? NO
3   Is there only one word that is not literal and non-compositional? YES, 'dog'
Therefore, it is NOT a core idiom, it is a ONCE.

A red herring
1   Is it compositional? NO
2   Are we able to recognise the untruth and pragmatically reinterpret it as a truth? NO
3   Is there only one word that is not literal and non-compositional? NO
Therefore, it is a core idiom.

Having established the criteria for deciding what were core idioms, the next step was to see how many there were and if possible to make a list of them.

Making a list of core idioms

The search for possible core idioms needed to be as exhaustive as possible to make sure that eventually a complete list could be made.

Idiom dictionaries

The first source of potential core idioms to be examined consisted of ten idiom dictionaries. The following dictionaries were thoroughly searched, beginning with the *Cambridge International Dictionary of Idioms.*

Cambridge International Dictionary of Idioms (McCarthy and Walter, 1998)
Collins COBUILD Dictionary of Idioms (Sinclair and Moon, 1995)
A Dictionary of Everyday Idioms (Manser, 1997)
English Idioms (Seidl and McCordie, 1988)
Longman Idioms Dictionary (Stern, 1998)
Oxford Dictionary of Idioms (Speake, 1999)
Oxford Dictionary of English Idioms (Cowie, Mackin and McCaig, 1993)
Oxford Idioms Dictionary for Learners of English (2001)

The Penguin Dictionary of English Idioms (Gulland and Hinds-Howell, 1986)
The Wordsworth Dictionary of Idioms (Kirkpatrick and Schwarz, 1993)

After applying the criteria of non-compositionality and non-figurativeness and distinguishing ONCEs to all the items in the *Cambridge International Dictionary of Idioms*, an inter-rater reliability check was carried out to make sure that the procedure was reliable and replicable. Another rater was separately given all the items listed on pages 80, 160, 240, 320, and 400 of the *Cambridge International Dictionary of Idioms* and was asked to identify those which qualified as core idioms. Time was spent giving some training by explaining how the definitions applied to a variety of items from other pages in the dictionary, and then seeing if the rater could apply the criteria to decide which of the test group qualified as core idioms. The rater was then left with the selected pages from the idiom dictionaries and asked to return the lists after completing them. The agreement reached was 91%.

The search of the Cambridge dictionary resulted in a total of 69 core idioms from the 5,782 entries in the dictionary. Then the remaining nine dictionaries were carefully examined and 32 additional core idioms were found, making a total of 101.

Other sources

The search was then widened to include five major English teaching books, seven vocabulary teaching books, 128 journal articles and chapters about idioms, several months of horoscopes from magazines and newspapers, sports reports, and several months of scripted TV programmes and unscripted programmes. Two additional core idioms (*row crosshanded, lay out in lavender*) were found in the journal articles bringing the total to 103. No additional core idioms were found in any of the other sources. This made us feel that the search of the dictionaries was unlikely to have missed any important core idioms.

The corpus search

Having established a comprehensive list of core idioms, the next step was to look at the occurrences of these idioms in a very large corpus to find out their frequency, their various forms, and whether they had literal equivalents. The British National Corpus was chosen for the search because it is large, is recent, has a wide range of texts, and includes both spoken and written data. The British National Corpus is approximately 100,000,000 running words long and contains both informative and imaginative prose. The SARA program which comes with the corpus was used for the searching.

The search was not straightforward because many of the idioms had various forms and all needed to be found. These different forms include differences in punctuation (especially hyphens and apostrophes), spelling, inflections, changes in word order, substitutions, and truncations. Here are two examples of the various forms that had to be searched for.

All in

A phrase such as *all in* is more commonly found as part of a larger phrase or clause (*all in all, all in the room, all in one place,* etc.), so it was necessary to do separate searches combined with the verb 'to be' (*am/is/are/was/were/been/being + all in*) plus the verbs 'look' (*look/looks/looked + all in*) and 'feel' (*feel/feels/felt/feeling + all in*). The phrase was also found with adverbs such as *about all in* or *absolutely all in*.

Frequency

Eighteen of the 103 core idioms did not occur at all in the British National Corpus. In one or two cases, *buy the farm*, this may be because the phrase represents American rather than British usage. Forty of these core idioms did not occur at all in the 10,000,000 word spoken section of the British National Corpus.

If the core idioms were treated in the same way as single words in a frequency count, none of them would get into a list of the 5,000 most frequent word families of English, and only seven would get into the top 7,000. Core idioms are not frequent and do not deserve classroom time. The frequency figures in Appendix 1 are based on the 100,000,000 token British National Corpus, so *by and large*, occurred 487 times in 100,000,000 running words, or just over 5 times per 1,000,000 words. As a comparison, 1,000,000 words is about 100 hours of continuous talking, or about three or four university text books, or about ten to twelve novels. Clearly this small number of opportunities for meeting the core idioms is not a good return for teaching. If the frequency of core idioms as a group is considered, the 103 core idioms had a total of 2260 occurrences in 100,000,000 running words. This works out as one occurrence of a core idiom per 34,000 running words, or roughly one core idiom every 100 pages of text. The BNC figures are a little inflated because they include some instances of citation, for example, all seven occurrences of *kick the bucket* are citations from discussions of idioms. Other forms however - for example *kicked the bucket* - occur in normal usage.

The search provided a wealth of data about each core idiom. The overall results of the search are as follows.

Red herring

Red herring is in all the dictionaries, but always in the singular form. A search for 'red herring' found 56 occurrences, while a search for the hyphenated 'red-herring' found another 2. However, when the plural 'red herrings' was searched for, an additional 26 occurrences were found. Therefore, all core idioms that could occur in the plural had to be searched for in this form as well.

Frozenness

Frozenness refers to the degree to which changes can occur in the form of a phrase. It is often assumed that most core idioms are frozen. In determining frozenness, the use or non-use of hyphens was ignored. *By and large* for example also occurs as *by-and-large* in two instances in the British National Corpus. Deliberate plays on words were also ignored. For example, *bats in the belfry* was found as a book title about clerical life entitled *Bare Butts in the Belfry*, and a story about publishers was entitled *Cats in the Belfry*. As a play on words can be made with almost any MWU, this was not considered a factor to determine frozenness. Therefore, *bats in the belfry* was still considered frozen.

Here are some of the variants of the non-frozen core idiom *pull someone's leg*. The data for considering frozenness was occurrence in the BNC. Core idioms occurring only once or not at all in the BNC could not be considered as more than one occurrence is needed to determine frozenness. Inflectional changes made an item non-frozen as did pronoun substitution.

> *Pull my blue leg, somebody's leg was being pulled, having his leg pulled, leg pulling, a leg pull, a leg puller, tugged my leg, yank somebody's leg, leg tugged/yanked.*

Take the piss can occur with all the verbal inflections of *take* – *taking the piss, takes the piss, took the piss,* and *piss-take, piss-takes* and *piss-taking* also occurred.

If hyphenation and other punctuation is ignored, then 27 of the 78 idioms occurring twice or more were frozen according to the examples gathered from the British National Corpus (see the italicised items in Appendix 1). This is a relatively small proportion showing that frozenness cannot be considered a typical characteristic of core idioms. This is consistent with the findings of Moon (1998).

Literal equivalents

A minor problem with recognizing core idioms and figuratives is realising that they are non-compositional and cannot be interpreted literally or as an idiom. During the search of the British National Corpus, a note was made of the number of times each MWU occurred in the same form but with a literal meaning. This information is of some value for teachers and learners, because learners can then be told that whenever they meet *by and large*, for example, it will always be a core idiom.

The potential for causing confusion to learners is clearly demonstrated with BNC examples of one core idiom, *Beat it!* (Go away!) which had only 12 core idiom occurrences but 78 literal occurrences including the following:

- Beat it (strike repeatedly – 30) –
  a. and whenever we came to a thicket or undergrowth we beat it with our sticks to stir up the prey.
  b. 'Tell me what day this is, Eleanor! You tell me or I'll beat it out of you!'
  c. I waded over to it, trying to drag the four hundred weight on its side, and when all else failed, I beat it with my fist...

- Beat it (overcome, surpass, win – 42)
  d. 'Nothing to beat it for shifting the wind!'
  e. Mark made 103 m.p.h. this time and Robinson just failed to beat it at 102 m.p.h., leaving the large crowd breathless!
  f. 'Can you beat it, Clive? They're actually running a bet on who I'm going to marry. Isn't life rich?'

- Beat it (stir vigorously – 6)
  g. 'Take a quart of green gooseberries boil them and put them thro' a sieve, take the white of 3 eggs, beat them to a Froth, put it to the Gooseberries and beat it both together till it looks white...
  h. 'Take as much lean of boiled ham as you please, and half the quantity of fat, cut it as thin as possible, beat it very fine with a mortar...'
  i. '...press out the juice, and draw the milk out of the cow's udder into it, sweeten it with a little sugar, and beat it well with birchen twigs till...'

Learners must first decide if the MWU has an idiomatic or a literal meaning, and if literal, which literal meaning is correct. They must also learn to look at the MWU and the words used with it. All BNC literal occurrences of *kick the bucket* included '*kick the bucket over*'. Only three items had more literal occurrences than core idiom occurrences, *beat it, a piece of cake, to boot.*

Although several items could possibly have literal occurrences, such as *the Big Apple* and *a French letter*. In fact they did not occur with such a literal meaning in the British National Corpus. Only 15 of the 103 core idiom forms had literal occurrences in the British National Corpus (see Appendix 1).

Teaching and learning multi-word units

In this section we will consider core idioms in comparison with other kinds of MWUs. From a learning perspective we can divide MWUs into three types – core idioms, figuratives, and literal sequences. Each of these categories requires a different approach to learning.

Core idioms

Core idioms (Grant and Bauer, 2004) cannot be understood from knowing the meaning of

their parts. They need to be learned as whole units. As we have seen, there are not many of these and only a few of them are reasonably frequent. The most useful core idioms need to be learned after the learners are in control of the high frequency words of the language and are well into expanding their low frequency vocabulary.

Deliberate learning of core idioms can be done by putting the core idioms on word cards with their translation on the back and deliberately learning them. The very low frequency of all but a few of the core idioms means that teaching would be largely a misdirection of both teaching effort and use of classroom time. There are many more useful and important words and phrases to learn compared with the majority of core idioms.

If, in spite of this, the teacher wants to teach core idioms, then there are useful things that can be done. If the origin of the core idiom can be related to its meaning, then telling the learners of this connection would be helpful. For example, *trip the light fantastic* comes from a poem by John Milton (*L'Allegro*) "Come and trip it as ye go. On the light fantastick toe" where *trip* means to walk lightly. In *go by the board(s)*, the *board* refers to the board of a ship and means that something has fallen overboard. Boers (2001) and Boers and Eyckmans (2004) have shown that understanding the etymology of an MWU can help learners remember it.

Some figuratives and a few core idioms make use of sound patterns like alliteration and rhyme (Boers and Lindstromberg, 2005). Pointing these out can help learning. For core idioms, this may work for *so and so, such and such, odds and sods, at sixes and sevens, all gas and gaiters*, and *Box and Cox*.

Because core idioms can be treated as a unit like a word, it is useful to see the grammatical patterns and contexts in which they occur. Seeing a concordance of contexts taken from a corpus can help with learning. Concordancing programs like WordSmith Tools and Monopro will be useful for getting examples. The use of concordances has been recommended by several teachers and researchers (Thurstun and Candlin, 1998; Stevens, 1991) and there is a small amount of research on their value and how they are best used (Cobb, 1997).

Figuratives

By way of contrast, let us look briefly at how the two other major categories of MWUs – figuratives and literal sequences – could be taught and learned.

Figuratives make up most of what people commonly call idioms. At first sight their parts do not make up the meaning of the whole unit.

> *kill two birds with one stone*
> *give the green light*

> *the apple of my eye*
> *as good as gold*

However, once the meaning of the MWU is known, it is possible to see how the parts go together to make the meaning known. *He is the apple of my eye* means "I am very proud of him", "I like him a lot". Knowing this, we can see that we could paraphrase the expression by saying "When I look at him with my eye(s), I see something that is very nice (like an apple)". For many figuratives, it is possible with a bit of skill and effort to interpret the phrase without having met it before, for example *it's as good as gold, this is just what the doctor ordered*.

Figuratives are unlikely to be directly translatable word by word into another language, and native speakers of English learn them through meeting them. People make up new figurative expressions which others manage to understand.

> *He is a sandwich short of a picnic.*

Pauwels (1995) points out that what we call figuratives (metaphors) can be placed on a scale of recoverability. At one extreme, approaching core idioms, are those where the original literal use cannot be recovered. That is the original literal meaning of the expression in the metaphor has fallen out of use and thus there is no support for working out the figurative use. Examples include *bandy words* (*bandy* was a game like tennis), and *the burden of his message*. At the other end of the scale of recoverability are expressions like *eat one's words* which are easily recoverable. Recoverability is increased if the literal meaning of the metaphor is still in use in the language, or if there are several current expressions "which provide clues for each other's interpretation" (p. 129). Geeraerts (1995) also notes that "the semantic specialization of idioms is a matter of degree" and that in some cases it is necessary to go back to a figurative expression to interpret a literal multi-word unit. The direction of recoverability is not always figurative to literal.

To deal with figuratives receptively, in listening and reading, learners need to have an interpreting strategy and will be greatly helped if they have actually met and learned the most useful figuratives. The strategy for interpreting figuratives is a commonsense one. The learner looks at the literal meaning of the MWU and then tries to see how this meaning needs to be reinterpreted to fit with the context in which it is being used. Here are two examples from the British National Corpus.

> *If we ever raise that hull we should be able to kill two birds with one stone: to determine the causes of both the explosions and this man's death*

> *You might as well kill two birds with one stone by doing and learning in parallel.*

Common literal sequences

There are many reasonably frequent MWUs that can be understood from the meaning of their parts. From the perspective of a particular group of second language learners, it is possible to divide literal sequences into two groups – those that have a word for word equivalent in the first language, and those that do not. From the viewpoint of a Korean learner of English, here are some that have first language equivalents.

High school, high income, high court

Once again from the perspective of a Korean learner of English here are literal sequences that do not have word for word equivalents in Korean.

High street, high water, high spirits

The reason for distinguishing these two groups is that those with first language equivalents need little learning as long as the words that make them up are already known. Kellerman (1985) however suggests that second language learners may be overcautious in anticipating parallels between L1 and L2. Those that do not have first language equivalents will not be difficult to understand, but will need to be noticed and remembered before they can be used in speaking and writing. Having deliberately noticed them will make later meetings with them easier to process. It is also possible to apply the L1 equivalent/no L1 equivalent distinction to core idioms and figuratives, although we would not expect to find many English core idioms with an L1 equivalent making use of the same parts.

Table 1 summarizes the types of MWUs discussed above.

Table 1. Types of MWUs, their transparency, and their learning requirements

types	Transparency	Learning needed
Core idioms	cannot be predicted or analysed	They must be learned as complete units
Figuratives	cannot be predicted and need to be interpreted	An interpreting strategy can be practised. They should be noticed and useful ones learned
Literals — no L1 equivalent	cannot be predicted but can be analysed	These need to be met and remembered
Literals — L1 equivalent	can be predicted and analysed	Deliberate learning is not needed except for early fluency development

what they mean when they talk about idioms. They are really grouping three different categories as one. Most of what are loosely called idioms are figuratives which require a different coping and learning strategy from core idioms. Second, true idioms (core idioms) are a very small and largely infrequent group of items. Core idioms should not be a major learning goal of a language learning programme. Figuratives however do deserve attention mainly because of their large number and because they can be dealt with receptively through a commonsense strategy. Learners need to recognise the different types of idioms when they encounter them and need to develop strategies for dealing with them.

This article has two important related messages. First, teachers and learners need to rethink

Boers, F. (2001). Remembering figurative idioms by hypothesising about their origin. *Prospect*, 16(3), 35-43.

Boers, F. and Eyckmans, J. (2004). Etymological elaboration as a strategy for learning figurative idioms. In P. Bogaards and B. Laufer, (Eds.), *Vocabulary in a Second Language: Selection, Acquisition and Testing*. Amsterdam/Philadelphia: John Benjamins: 53-78.

Boers, F. and Lindstromberg, S. (2005). Finding ways to make phrase-learning feasible: The mnemonic effect of alliteration. *System* 33, 225-238.

Cobb, T. (1997). Is there any measurable learning from hands-on concordancing? *System* 25, 3: 301-315.

Geeraerts, D. (1995). Specialization and reinterpretation in idioms. In M. Everaert et al (Eds.) *Idioms: Structural and Psychological Perspectives*. Hillsdale: Lawrence Erlbaum Associates.

Grant, L. and Bauer, L. (2004). Criteria for re-defining idioms: Are we barking up the wrong tree? *Applied Linguistics* 25(1), 38-61.

Kellerman, E. (1985). If at first you do succeed .. In S.M. Gass and C.G. Madden (Eds.) *Input in Second Language Acquisition*. Rowley, Mass: Newbury House: 345-353.

Moon, R. (1998). *Fixed Expressions and Idioms in English: A Corpus-Based Approach*. Oxford: Clarendon Press.

Pauwels, P. (1995). Levels of metaphorization: the case of *put*. In L. Goosens et al (Eds.) *By Word of Mouth*. Amsterdam: John Benjamins.

Stevens, V. (1991). Classroom concordancing: vocabulary materials derived from relevant, authentic text. *English for Specific Purposes* 10: 35-46.

Thurstun, J. and Candlin, C.N. (1998). Concordancing and the teaching of the vocabulary of academic English. *English For Specific Purposes* 17, 3: 267-280.

Idiom dictionaries

Cowie, A. P., Mackin, R., & McCaig, I. R. (1993). *Oxford Dictionary of English Idioms*. Oxford: Oxford University Press.

Gulland, D.M., & Hinds-Howell, D. (1986). *The Penguin Dictionary of English Idioms*. Harmondsworth: Penguin Books.

Kirkpatrick, E. M., & Schwarz, C. M. (Eds.). (1993). *The Wordsworth Dictionary of Idioms*. Ware, Hertfordshire: Wordsworth Editions Ltd.

Manser, M. H. (1997). *A Dictionary of Everyday Idioms* (Second ed.). London: Prentice Hall Europe ELT.

McCarthy, M. J. & Walter, E. (1997). *Cambridge International Dictionary of Phrasal Verbs*. Cambridge: Cambridge University Press.

Oxford Idioms Dictionary for Learners of English. (2001). Oxford: Oxford University Press.

Seidl, J., & McMordie, W. (1988). *English Idioms* (Fifth ed.). Oxford: Oxford University Press.

Sinclair, J., & Moon, R. (1995). *Collins COBUILD Dictionary of Idioms*. London: Harper Collins Publishers.

Speake, J. (1999). *The Oxford Dictionary of Idioms*. New York: Oxford University Press.

Stern, K. (1998). *Longman Idioms Dictionary*. Harlow, Essex: Longman.

Appendix 1: Core idioms and their frequency in the 100,000,000 word British National Corpus

(Italicised items are frozen; those followed by an asterisk* have literal equivalents in the BNC; underlined items have an unusual grammatical form.)

Idiom	Freq
by and large	487
so and so	327
such and such	196
out of hand	141
take the piss	137
and what have you	136
serve sb right	101
take sb to task	92
red herring	87
(be) beside sbself*	72
out and out *	72
take the mickey	71
pull sb's leg*	60
touch and go	53
the Big Apple	52
cut no ice with sb	50
come a cropper	49
put your foot in it	48
an axe to grind	47
make no bones about it	44
a piece of cake*	43
a white elephant	43
(all) of a piece	38
Uncle Sam *	35
and what not	33
go by the boards	32
so long!	24
eat your heart out	23
not hold a candle to	23
off the wall	22
pull the other one	19
trip the light fantastic	18
cut a dash	18
to a T	18
by and by	17

Idiom	Freq
push the boat out*	16
gird your loins	16
hell for leather	15
cook a hoop	15
Bob's your uncle	14
chew the fat*	14
cold turkey*	14
kick the bucket*	13
the cat's whiskers	13
give sb the bird*	13
(be) all in*	13
a French letter	12
not cut the mustard	12
odds and sods	12
Beat it!*	12
butter wouldn't melt in sb's mouth	11
sweet Fanny Adams	10
to boot*	10
knock into a cocked hat	9
cook sb's goose	8
do your nut	8
Old Glory	8
cut up rough	8
pop your clogs	8
bang to rights	8
the bee's knees	7
at sixes and sevens	7
like nobody's business	7
cut a rug	6
off your face	6
get your rocks off	5
have a bone to pick with sb	5
nip and tuck	4
shoot the breeze	4
(raining) cats and dogs*	3

APPENDIX

heap coals of fire on sb head	3
swing the lead	3
have bats in the belfry	2
sell sb a bill of goods	2
splice the mainbrace	2
a fine kettle of fish	2
set sb by the ears	2
take the rise out of sb	2
lay an egg	1
how's your father	1
make sth out of whole cloth	1
all gas and gaiters	1
make your number	1
half seas over	1
hop the twig	1
Box and Cox	0
buy the farm*	0
come the raw prawn	0
go bung	0
the ghost walks	0
part brass rags with	0
bust sb's chops	0
the bitch goddess	0
row crosshanded	0
clean sb's clock	0
the goose hangs high	0
tie one on	0
eat crow	0
lay out in lavender	0
over the transom	0
to a fare-thee-well	0
peg it	0
wear the green willow	0