Radoslav Pavlík

On Some Basic Characteristics of Idiomatic Expressions

Radoslav Pavlík

On Some Basic Characteristics of Idiomatic Expressions

1 An Introduction

One of the properties of language is its creativity. This creative dimension of language, emphasized by Noam Chomsky, rests upon the fact that knowing the rules of grammar we are able to produce and understand an infinite number of new expressions (Chomsky 1975: 10–12; 2002: 13–25). However, our use of language is not governed by this principle exclusively. It often happens that students of English produce phrases and sentences in agreement with grammatical rules, but, much to their disappointment, they are told by the native speaker that a particular phrase or sentence, although grammatically well-formed, just does not sound right. In other words, even though such syntagmatic constructions are not ungrammatical, native speakers would never use them (Allerton 1984: 39). ‘Gaining full command of a new language requires the learner to become sensitive to the native speaker’s preferences for certain sequences of words over others that might appear just as possible’ (Wray 2000: 463). This reminds us of the fact that knowing the lexicon and the rules of grammar is not enough to produce natural native-like language. Anyone who tries to learn a language just by memorizing the rules of grammar and individual words is bound to fail (Milkov 2000: 88). This is because many expressions we produce or hear are not created as brand new, but exist in language as prefabricated chunks of words. Such expressions are familiar and conventional, and they are characterized as idiomatic (Lennon 1998:14; Fernando 1996: 31). It has been estimated that a large proportion of language (by some estimates up to 80%!) is idiomatic, which means that much of what we produce is in some way prefabricated and formulaic, and a great deal of discourse is more or less routinized (Halliday 1978: 4; Nattinger 1988: 76). Rather than being peripheral to the ‘core’ of a language, it is possible to argue that such idiomatic expressions are the core (Taylor 2002: 541).

Psycholinguistic evidence suggests, that such pre-constructed expressions enable our speech to be more fluent and economical. In our everyday lives, we often function in set routines, encounter many recurring situations and familiar settings, and interact socially in more or less set ways. All this is reflected in language in the form of ready-made and relatively stable phrases stored in our memories, and available to us whenever we need to express ourselves concisely and pithily (Lennon
1998: 16; Singleton 2000: 55). In short, when it comes to language, humans prefer using memory and routine possibilities. ‘If this proves inadequate, they turn to computation’ (Aitchison 1987: 3). Although there appears to be a general agreement as to the existence and wide use of idiomatic expressions, no consensus has yet been reached in terms of their naming or precise categorisation. The following are some examples of the names given to this linguistic unit:

collocation, idiom, phraseme, lexical phrase, prefabricated multiword expression, recurrent word sequence, recurrent word-chain, multiword unit, multiword chunk, fixed combination, set expression, lexicalized phrase, stock phrase, phrasal lexeme, semi-preconstructed phrase, conventional word association, cohesive word-cluster, cohesive combination, multiword string, conventional syntagma, preassembled phrase, preassembled chunks, etc.

The apparent plethora of expressions is due to many factors, and it would be pointless to give reasons for such an unfortunate state of affairs in a paper of this sort. Instead, we shall look for a way out of this terminological havoc by focusing on what most of the above expressions seem to have in common.

First of all, the category of linguistic units we are concerned with here has traditionally subsumed expressions with both transparent and non-transparent meaning. This is the issue dealing with the semantic properties of such expressions, and we will discuss it in later sections.

Secondly, most of these expressions denote a complex linguistic form, i.e. a form consisting of more than one word. This is indicated by the use of such lexemes as phrase, sequence, chain, chunk, combination, string, etc. It should be remembered that the definition of the very notion of the ‘word’ is not unproblematic, but limitations of space prevent us from discussing these issues here.

Since the phrase ‘multiword expression’ seems to be general enough to cover all cases of word combinations (transparent or non-transparent in meaning), we will use it as a part of our definition in this paper.

Finally, all of these expressions are idiomatic, which means that at the moment of production they are not composed on-line, but they are retrieved as prefabricated wholes. The words like idiom, prefabricated, recurrent, fixed, conventional, preassembled, etc. serve as indicators of idiomaticity of such expressions. Surprisingly, the word idiomatic is not used very often, although it could serve as an umbrella term for all of them.

In light of these facts, we will name this category of linguistic forms ‘idiomatic multiword expressions’, because it captures the essence of all the above expressions in an appropriate and sufficiently general manner. The discipline studying such idiomatic expressions will be referred to as idiomatology.

From what has been said it follows that our language production is governed by two main principles (Sinclair 1991):

(a) The idiom principle or (in line with our definition) the principle of idiomaticity, which states that part of language is idiomatic – consisting of prefabricated multiword expressions stored and retrieved as wholes rather than being generated by rules of grammar, even though they are analyzable into segments (Wray 2000: 480), e.g. heavy rain, be afraid of, as far as ... is concerned, it’s not what you think, happy birthday, what’s up?, red herring, bark up the wrong tree, jump down sb’s throat, etc.

(b) The open choice principle, according to which we can freely combine individual lexemes by means of a set of grammar rules, e.g.
These two principles complement each other, and they are present in all natural languages.

2 Formal Characteristics of Idiomatic Multiword Expressions

From a formal point of view, idiomatic multiword expressions can be characterized in terms of their two main features: multiword character and idiomaticity.

1. Multiword character
Expressions of this sort consist of more than one word and do not exceed the bounds of the sentence (Kavka 2003: 167), e.g. to believe in, on and on, bark up the wrong tree, appearances are deceptive, etc. Multiword expressions should be distinguished from such forms as compounds, which are usually taken as one single word (although sometimes problems might arise with compounds written as two separate elements). The concept of a multiword expression given in this paper is thus different from other theories, which include compounds in this category, e.g. Filipec and Čermák (1985), Fernando (1996), inter alia.

2. Idiomaticity
The term idiomaticity refers to all those features of language that are connected with prefabricated ready-made expressions that are, at the moment of production, not generated by the rules of grammar, but stored and retrieved from memory as wholes (Lennon 1998: 16). We will distinguish between two aspects of idiomaticity: collocability and conventionality.

A. Collocability (Lexical co-occurrence)
Certain words, in the stream of speech, are frequently found in the company of other words. Put differently, in particular contexts, some lexical items are likely to be accompanied by a limited range of other lexical items (Firth 1957; Halliday 1966)
The term collocability is usually used in connection with this statistically significant co-occurrence of words, and it is generally determined by means of a computer analysis of large corpora (collections of spoken and written texts). In principle, it is a matter of counting the number of times a particular word occurs in the company of another word. If words appear together much more often than by chance, we speak of collocability (Manning and Schütze 1999: 143; Inkpen and Hirst 2002: 68). Collocability can be of different degrees. At one end, we have multiword expressions the elements of which occur together very frequently (they are completely predictable), e.g. rancid only collocates with butter and oil (rancid butter, rancid oil), and spick can only combine with and span (spick and span). At the other end of the spectrum, there are multiword expressions consisting of words which can be combined with a whole range of other words, but these expressions are nonetheless predictable, because the range of combinatorial possibilities of the words constituting the expression is not unlimited, e.g. see a movie, see a doctor, see reason, etc. It should be noted that although the term collocability is used mainly in connection with collocations, ‘... this characteristic is typical of idioms too’ (Barkema 1996: 129), and we will therefore use it as a general attribute of all multiword expressions.
B. Conventionality (Institutionalization, Familiarity, Prefabrication, Formulaicity)
The main principle of generative grammars is that lexical items can be combined into standard sequences by means of grammatical rules, e.g. *a big house*, *to buy milk*, etc. Many such constructions, whether literal or metaphorical, become institutionalized over time, which means that they gradually become more current and familiar within a language community. Such conventional multiword expressions are considered prefabricated in the sense that they are felt by a native speaker as something compact and ready-made, stored in our memories for future retrieval. Conventionality differs from collocability in that it refers to the intuitive knowledge of a language user, whereas collocability is defined in this paper as a statistically significant co-occurrence of words determined by means of a computer analysis of large corpora of texts. Conventionality is about linguistic introspection. This means that when a speaker is given a particular word (usually in some context), he or she can supply a list of words which typically combine with the given word (Greenbaum 1974: 83; Herbst 1996: 379–381). For example, when a native speaker is asked to fill in an appropriate word in the construction ‘______ rain’, such adjectives as *heavy*, *driving*, etc. will automatically spring to mind. Another way of finding out whether a combination of words is conventional is to ask a native speaker whether a particular expression sounds natural. For instance, the expressions *heavy rain* and *rolling stone* will be judged as natural and appropriate by all native speakers of English, whereas the combinations *strong rain* and *rolling rock* will be judged as odd and unnatural. It goes without saying that collocability and conventionality are related. When a multiword expression is conventional, it is likely that its parts will be found to collocate, even though such collocability may have different degrees of restrictedness (Aisenstadt 1981; Howarth 1998). At the same time, a multiword expression whose parts collocate with the probability higher than chance is likely to be conventional, although this may not always be the case, e.g. such multiword sequences as of the, in the, for the, a few, and then, and that, does not, and he was, and there was, etc. (Manning and Schütze 1999: 144; Stubbs and Barth 2003: 70) are statistically highly frequent, but they are not normally felt to be conventional.

Before concluding this section, a brief comment on the terms *fixity*, *fixedness*, *stability*, and the like is in order. Currently, there are a number of linguists who work with the notion of stability, and they consider this characteristic to be one of the defining features of idiomatic multiword expressions, mainly, the idiomatic multiword expressions with non-transparent meaning – idioms (Mlacek 1984; Cowie 1988; Fernando 1996, inter alia). We do not consider this feature to be significant enough to be included in the definition. The reason for this is twofold. First of all, it has not been stated explicitly so far what exactly is meant by stability. For example, are we talking about the stability of the relation ‘form–form’? Because if we are, this is already covered by the notion of collocability and conventionality. Furthermore, we can talk of the stability of the relation ‘form–meaning’. That is, how one meaning is expressed by one or more forms. If a particular concept is expressed by one form only, than the stability of the ‘form–meaning’ relation is absolute. What is interesting with respect to this approach is that the stability status of many collocations is much higher than that of idioms. For instance, the concept *WHITE WINE* can be expressed by means of the collocation (*white wine*) or a free combination (*an alcoholic drink of light colour made from grapes*). Idioms, conversely, have always their non-metaphorical counterparts, so there is usually at least one more form in comparison to a collocation, e.g. *kick the bucket* vs. *die*, *stop living and become dead*, *pass away*, etc. Incidentally, the confusion surrounding the definition of the term *restrictedness* (of collocations) is largely due to ignoring the distinction between the ‘form–form’ and ‘form–meaning’ relations. There is yet another type of stability – the stability of the overall form in terms of word order, inflectional variation, insertion of elements not considered to be part of the expression, etc. Unfortunately, it is not possible to address these questions in a paper of this sort.

Secondly, it has been emphasized repeatedly that idiomatic forms are by no means as fixed as some dictionary inventories appear to suggest, and this concerns both institutionalized variation and
creative modification (Moon 1998: 81, 96). Finally, some researchers do not work with the notion of stability at all (Titone and Connine 1994).

It is evident from this brief discussion that the problem of the so called ‘stability’ deserves more attention, but a thorough discussion of these issues lies well outside the scope of this paper.

3 Semantic Characteristics of Idiomatic Multiword Expressions

A number of idiomatic multiword expressions are to some extent opaque. That is to say, their meaning is frequently not deducible from the meaning of the items they are composed of. This, however, is a simplification of the facts, because in reality we can distinguish among many different degrees of opacity (non-transparency of meaning), ranging from expressions that are fully opaque to expressions that are completely transparent (Mlacek 1984: 69–74; Fernando 1996: 35–36; Cowie 1998: 4–7). Within the framework of this handbook, we will set up the following degrees of opacity/transparency of multiword expressions:

1. Opaque multiword expressions
These expressions are considered to be completely non-transparent in the sense that there is absolutely no connection (synchronically) between the meanings of the individual components of an expression and its meaning as a whole (Titone and Connine 1994: 247; Kvetko 2003: 32–33). For example, the multiword expression *fat cat* is not to be taken literally as an ‘obese feline’, but it is used to denote somebody who is rich and powerful. These opaque expressions are also called pure idioms or phraseological fusions (Vinogradov 1947; Cowie 1998: 4–7).

2. Semi-opaque (Figurative) multiword expressions
There are types of multiword expressions that have originated as metaphors, but through their constant use they lost much of their characteristic flavour, and interpreting them no longer requires the use of the metaphorical strategy. They are viewed as ready-made prefabricated chunks of words and understood without recourse to their literal meaning. At the same time, however, the connection between the meaning of the components and the expression as a whole is not completely lost (Cruse 1986: 42). In other words, we can see a link between the literal and transferred meaning of such expressions. These expressions are referred to as dead metaphors, phraseological unities, figurative idioms, or semi-opaque idioms (Vinogradov 1947; Cowie 1998: 5; Kvetko 2003: 33). Examples of semi-opaque multiword expressions are copious: *roll up one's sleeves, sweeten the pill, put one's cards on the table*, etc. It should be pointed out, however, that the boundary between opaque and semi-opaque expressions is not clear cut, but depends on the cultural experience of an individual.

3. Semi-transparent multiword expressions
In this type of multiword expressions, one word is used in its literal sense, while the meaning of the other word(s) is more or less opaque. Such expressions are known as phraseological combinations or semi-idioms (Vinogradov 1947; Cowie 1998: 5). Semi-transparent multiword expressions form the dividing line between idioms and collocations, since they combine both opaque and transparent items.

It has been pointed out repeatedly that this group of idiomatic expressions is difficult to separate from transparent idiomatic expressions usually called collocations. The main problem lies in the degree of transparency/opacity of the (partially or fully) non-transparent element. In other words, since this element (word) can be non-transparent to various degrees, it is difficult to determine objectively where the cut-off point should be placed dividing (relatively) non-transparent idioms from (relatively) transparent collocations.

The solution we offer in this paper is based these criteria:
A multiword expression will be classified as a semi-transparent idiom when
(a) one or more words are used in their literal sense and at least one word is used in a non-literal (transferred) sense,
(b) the transferred sense of the word(s) must be considered to be homonymous in relation to the literal sense of that word, i.e. the form of that word has two different meanings,
(c) words with this new meaning are unique and found only in one or two expressions.

Examples of semi-transparent multiword expressions are such combinations as foot the bill (= pay the bill) or it’s raining cats and dogs (= it’s raining heavily).

As we can see, the two meanings of the words foot and cats and dogs are completely different, and, at the same time, words with these new meanings are unique, e.g. we cannot use the word foot in the sense ‘to pay’ in any other expression. The same goes for cats and dogs (meaning heavily/hard).

These three categories we have defined so far (opaque, semi-opaque, and semi-transparent) will be, within the framework of this paper, subsumed under the heading of phraseology (idiomology) – the discipline dealing with multiword expressions with non-transparent meaning – idioms. It should be noted, however, that the term phraseology is often used as a cover term for all idiomatic multiword expressions (e.g. Melčuk 1998; Altenberg 1998; inter alia), in which case it is identical to what we call here idiomatology.

4. Quasi-transparent multiword expressions
These multiword expressions are no longer considered to be idioms, because all the component parts of these expressions are more or less transparent in their meaning. This group of expressions is similar to the previous one in that one or more words are used in their literal sense and at least one word is used in a non-literal (transferred) sense. In other words, the two groups share the criterion (a). The two remaining criteria, however, are different:
(b) the transferred sense of one of the components of the expression must be considered to be polysemous in relation to the literal sense of that component, i.e. the form of that word has two different, but (synchronically) related senses (not different meanings).
(c) words with this new sense are frequent and found in many other expressions.

Such expressions belong to the category of quasi-transparent collocations. For example, the collocation black coffee is treated as quasi-transparent, because the meaning of the word black is not to be taken literally (of black colour), but it refers to a dark colour of various shades, or something which appears to be black. This new transferred meaning, however, is seen as polysemous rather than homonymous in relation to the literal meaning. At the same time, the word with this transferred meaning is not unique and restricted to the collocation black coffee only – it is found in many other expressions: black man, black tea, black cherry, black clouds, black pepper, etc.

5. Transparent multiword expressions
Expressions of this sort are considered to be fully transparent, which means that the meaning of the whole can be deduced from the meanings of the individual components of an expression. In other words, these expressions are co-occurrences of ‘... two or more words whose meaning can be inferred from the parts, but will become less acceptable when one of the elements is replaced by a synonym’ (Shei and Pain 2000: 167), e.g. a flock of geese, safe from, to make a mistake, etc. However, they differ from (relatively) free combinations of words (which are also fully transparent in their meanings) in that they are idiomatic, i.e. they are conventional and their elements co-occur
One last comment is needed on the term ‘free combination’. Although, in principle, it is possible to speak about free combinations of words, it should be remembered that such term is, *sensu stricto*, misleading. That is to say, ‘... all words have a limited potential to combine with other words and there is no such thing as a word without any collocational restriction’ (Čermák 2001: 2). For practical purposes, however, we see no harm in using the term *(relatively)* free combinations, because, from the language user’s perspective, such combinations may indeed appear to be free (considering their high degree of collocability).

The following table gives a schematic overview of the main categories and characteristics of idiomatic multiword expressions discussed in this paper.

<table>
<thead>
<tr>
<th>Idiomaticity</th>
<th>Type of multiword syntagm</th>
<th>Degree of opacity/ transparency</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-idiomatic</td>
<td>Free combination</td>
<td>Fully transparent</td>
<td>white colour</td>
</tr>
<tr>
<td>Idiomatic</td>
<td>Non-phraseological expression</td>
<td>Fully transparent</td>
<td>white cloud</td>
</tr>
<tr>
<td></td>
<td>Collocation severed</td>
<td>Quasi-transparent</td>
<td>white meat</td>
</tr>
<tr>
<td></td>
<td>Idiom</td>
<td>Semi-transparent</td>
<td>white lie</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Semi-opaque</td>
<td>white flag</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Opaque</td>
<td>white elephant</td>
</tr>
</tbody>
</table>

4 Conclusion

The production of every natural language is governed by two complementary principles: the principle of idiomaticity and the open choice principle. The first of these principles gives rise to a number of idiomatic expressions which exist in language as prefabricated chunks of words, ready for immediate retrieval. In other words, such multiword expressions are stored and retrieved as wholes rather than being generated by rules of grammar, even though they are analyzable into segments. The awareness of the existence of such idiomatic expressions is extremely important for the theories of language teaching and learning, since knowing the lexicon and the rules of grammar is by no means sufficient for the production of natural native-like language. The main characteristic features of *multiword idiomatic expressions* are, as the term indicates, multiword character and idiomaticity. From the point of view of the transparency/opacity of their meaning, they have been divided into five groups: opaque, semi-opaque, semi-transparent, quasi-transparent, and fully transparent. The first three groups are subsumed under the heading of *idiom*, while the remaining two belong to the domain of what has traditionally been termed *collocation*.

**References**


