1. Collocations

1.1. Introduction

The first recorded mention of the term collocation in a distinctly linguistic context listed under the entry for collocation in the second edition of the Oxford English Dictionary (OED 2nd) dates back to a quotation by Harris of the year 1750: “The accusative… in modern languages… being subsequent to its verb, in the collocation of the words.

Etymologically, the term goes back to the Latin collocate-us, the past participle of collocäre ‘to place side by side’ from col- (con-) ‘together’ + locäre ‘to place’.

In the quotation by Harris, collocation is used in a sense that is now commonly covered by the closely related term colligation, i.e. the grammatical juxtaposition of words in sentences. No mention is made of the strongly lexical character that is nowadays associated with the concept of collocation over and above the grammatical relations holding between their constituents (Sabine Bartsch 2004: 28). The term colligation denotes merely a grammatical relation between lexical items, whereas the concept of collocation entails a grammatical relation between lexical items (a direct syntactic relation between the constituents) as well as particular lexical co-selection constraints on the choice of lexical items that can co-occur.

Among the first linguists to base a theory of meaning on the notion of “meaning by collocation” is J. R. Firth (1957) who is commonly credited with systematically introducing the concept of collocation into linguistic theory.

Publications by Trager and a book by Palmer (1938) using the term collocation in its subtitle suggest that Firth might not have been the first to use the term in the sense intended here (idem: 30). There is, indeed, reason to believe that Otto Jespersen (1917, according to Mackin 1978: 149-165) might have been the first to use the term collocation in the context of lexical co-selection:

“Little and few are also incomplete negatives: note the frequent collocation with no: there is little or no danger”.

Yet, it is Firth who is commonly associated with the term in linguistics and who developed his models, based on the notion of ‘meaning by collocation’.

1.2. Definition of collocation

Despite various attempts at capturing the essence of collocations, the concept is still a notoriously difficult one in linguistics. Beyond a very general characterization of collocations as frequently recurrent co-occurrences of lexical
items, the structures and regularities behind this phenomenon are still largely unknown (Bartsch 2004: 65).

The term *collocation* is used by different linguists, to refer to what are often very different combinations of word forms. As used in most studies, it refers to combinations of two lexical items which make an isolable semantic contribution, belong to different word classes and show a restricted range.

1.3. Defining Criteria

A first criterion (two lexical items, not grammatical ones) separates lexical combinations (collocations in a wider sense) from grammatical combinations.

One can note that the term collocation is sometimes used to refer to combinations of lexical items with grammatical items or grammatical constructions, like *do in, put up with, under consideration, and finish/stop + -ing* construction (see Benson et al. 1986: IX-XXIII, who distinguish between lexical and grammatical collocations). The third criterion (independent meaning of constituents) marks non-idiomatic combinations off from idiomatic class and restricted range allow us to distinguish collocations in the narrow sense from other lexical, non-idiomatic combinations.

*Meaning.* Each lexeme makes an independent contribution to the meaning of the whole collocation, or as Cruse (1986: 40) puts it, “each lexical constituent is also a semantic constituent”. This means that the constituents of a collocation can have special meanings which are restricted to one particular collocation.

If we consider the meaning of *white* in *white paint* or *white snow* as the most frequent meaning, then *white* in *white coffee* (‘with milk’), *white currant*, *white grape*, or *white wine* are different, while the meanings of the adjective in *white lie* (‘harmless’), *white night* (‘sleepless’) or *white sale* (‘of house’) are more opaque and are linked closely to the respective noun.

In *white horses* (‘white-topped waves’) and *white coal* (‘water as energy source’), it is not only the adjectives, but also the nouns that show special, metaphorical meanings. One might be tempted to call them idioms if one were not aware of the metaphor (Gramley and Pätzold 1992: 62).

Most research has concentrated on combinations of lexemes like *river-rise, agree-entirely, or fine-heavy*. These are indeed collocations from the semantic point of view because each constituent has independent meaning. These are in contrast with idioms like *paint the town red or fly off the handle*, where it is difficult to say what the individual constituents contribute to the meaning of the combination as a whole. In between these two clear-cut cases are, however many expressions like *It's raining cats and dogs*, often mentioned as a typical English idiom, do not qualify for full idiomatic status because at least one constituent is independently meaningful (*rain*) while the other is idiomatic (*cats and dogs ‘heavily’*). These intermediate cases are sometimes called *partial or unilateral idioms* (idem: 62).

*Word Classes.* This criterion refers to the fact that lexemes belong to different word classes. Examples are *demand-meet* (N-V), *hopes-high* (N-Adj.),
Semantic Transparency And Opacity In Fixed Expressions

and apologize-profusely (V-Adv.). This sets them apart from a series of lexemes of the same word class which contribute to the textual cohesion.

Range. The last criterion, restricted range, helps to set up different classes within the two-lexeme type of combination, namely, free combinations as opposed to collocations. As used by Firth, the term collocation refers to any combination of two lexical items. Items that are not closely related to others enter into free combinations (other scholars use the term unrestricted collocations) while the closer associations between lexemes are called collocations (or restricted collocations). Lexemes that belong to the core vocabulary of English are typically found in free combinations (cheap, expensive, great, long, large, interesting, old, young; buy, give, go, run, talk etc.). The number of combinations formed with them is enormous. When one talks about the number of lexemes (or collocates) that occur together (or collocate) with the lexeme under discussion (the node), one is concerned with that lexeme’s range (Jones and Sinclair 1974).

It is common to express the difference between free combinations and collocations in terms of their ranges: the range of lexemes in collocations is smaller than that of lexemes in free combinations. Members of a range are often characterized by shared semantic features, as in the case of the verb meet, which collocates, for example, with need, condition and requirement on the one hand, and bus, plane and train on the other (Gramley and Pätzold 1992: 63).

In fixed (unique, frozen) collocations lexemes have only one collocate: ajar with door, auburn and hair, nod and head, shrug and shoulders. There are not many frozen collocations, and it would be rash to say that the range of a certain lexeme is limited to one collocate, only as lexemes can extend their range and individual usage varies.

It is hard to distinguish between collocations and free combinations. There is no empirical research based on corpora large enough to allow the criterion of repeated co-occurrence to be tested conclusively and convincingly (Gramley and Pätzold 1992: 64). Jones and Sinclair (1974) base their investigations on a corpus of 135,000 words of spoken text; Johansson and Hofland use the LOB (Lund – Oslo – Bergen) Corpus of about one million words of written text, but remark: “we would like to have a distributional lexicon… summarizing the grammatical and collocational properties of words. This requires a far larger material (Johansson and Hofland 1989: 14).

There is also a conflict between the criteria for collocations, namely close association, repeated co-occurrence and limited number of collocates (Benson et al. 1986: XXIV). As long as the defining criteria are in conflict with each other there is no easy solution in sight to the problem of distinguishing between collocations and free combinations (Gramley and Pätzold 1992: 65).

Sabine Bartsch (2004: 68) notices that there can be little dispute over the minimal criteria present in virtually all definitions of collocations in the literature, but beyond a fairly vague common core of defining criteria, the literature abounds in different views regarding collocations.
Bartsch proposes two main types of criteria for collocations: quantitative and qualitative criteria.

**Qualitative criteria. Direct Syntactic relation**

Rösel (1995: 172 ff, according to Bartsch 2004: 71) introduces the requirement that in order for a word combination to deserve collocation status, the constituents must be in a direct syntactic relation with each other. This criterion of a direct syntactic relation between the constituents is crucial as it rules out as collocations many words co-occurring within the same sentence because they belong to a certain word field. If it assumed that there is a direct relation of mutual or directional influence between constituents of a collocation, this influence is likely to hold at all levels of linguistic structure. A syntactic relation between words in a word combination indicates a potential semantic relation. This relation may become relatively stable and established in the language when the word combination occurs frequently in always the same constellation and meaning. Thus, a direct syntactic relation between the constituents of a collocation is taken to be a defining criterion of a collocation.

**Meaning and collocation.**

Bartsch (2004: 71) considers that the semantics of a collocation has to be studied under four related aspects:

- the contribution of the individual constituents to the meaning of the collocation;
- the impact of occurrence within a collocation on the meaning of the constituents;
- the semantic analyzability of the collocation (transparency of meaning);
- the impact of the collocation on the wider context and vice versa.

Benson et al. (1986: 254) hold that collocations are transparent in meaning, as opposed to idioms which are held to be fully opaque. They classify different types of multi-word expressions along a continuum of decreasing semantic transparency. However, semantic transparency and opacity are notions that are very hard to capture in operational terms. Not all collocations are equally transparent in meaning.

When characterizing collocations in terms of their semantic transparency there appear to be four possibilities (Bartsch 2004: 72-75):

1) All constituents of the collocation contribute an aspect of their transparent meaning; the collocation remains semantically fully transparent in the sense that its meaning is constituted of overt realisations of one of the potential senses of each of its constituents.

2) At least one of the constituents of the collocation does not contribute lexical meaning:

3) One constituent may be delexicalised, i.e. lose all or part of its independent meaning, while the other constituent(s) retain their full lexical meaning.

4) The collocation remains superficially transparent, but carries an additional element of meaning that is not overtly expressed by any of its
Semantic Transparency And Opacity In Fixed Expressions

constituents: e.g. collocations as carries of cultural stereotypes such as the collocation consenting adult which, beyond denoting the general fact that an adult is perceived to be in a position to give his/her consent to particular events or actions, the collocation carries the additional meaning of a male person legally allowed to engage in homosexual relationships. This type of additional opaque information is rooted in the cultural and social background of a British English linguistic community.

A fourth type comprises partly opaque collocations in which (at least) one of the constituents acquires a collocation – specific meaning which it does not have outside this particular word combination: stiff drink / breeze / challenge ‘strong, intense’; a hung parliament ‘without a working majority’; run for an office / president ‘be (presidential) candidate’.

Such collocations represent metaphorical extensions of the meanings of their constituents, often restricted to a specific domain or situational context.

Thus, there are degrees of semantic transparency in collocations. Collocations are therefore often perceived as being semantically transparent. Yet, this cannot be taken to imply that their meaning is always a compositional function of the meanings of their constituents. The question of the compositionality of word meanings is a crucial one in linguistics. In the context of collocations it is difficult to establish which of the constituents contributes which proportion of the meaning of the collocation.

Bartsch’s (2004: 75) study takes reduced semantic transparency of a word combination as an indication of a collocation while fully recognizing that the semantic aspects of collocations are hard to capture except by studying collocations within their wider textual and domain context. A logical consequence of this requirement is the methodological principle taken in her study that collocations can only be studied fruitfully based on a corpus of authentic language.

1.4. Morpho-syntactic features of collocations

Collocations show various degrees of fixedness (Gramley and Pätzold 1992: 65).

Morphology: in some collocations the adverbs are not formally marked by the -ly morpheme: swear-blind; drunk-blind; forget-clean.

Substitutability: lexemes can often be replaced by close synonyms, for example: hardened criminal is found side by side with confirmed criminal, though *hardened burglar or *hardened murderer are not found. Conflict collocates with end and resolve, though not with *finish and solve.

Additions: additions, most often pre- or post-modifying nouns, are normal.

Deletion: although deletions are not impossible they are much rarer than additions: I have not got the faintest, foggiest (idea).

Displacement: personal pronouns may replace the actual collocational items.

Separability: some collocations cannot be separated, for example foot the bill and curry favour, which are called bound collocations (Cruse 1986: 41).
Distribution: finally the distribution or word order of the constituent lexemes in collocations is relatively free: they met their demands; their demands, which were not met completely.

Syntactic transformations are thus possible and do not change, or destroy the meaning of collocations.

Collocations are less fixed than other types of expression. This low degree of formal fixedness in combination with the composite (isolable) semantic structure can also explain why collocations are rarely, if ever, the object of word play. Playful variation can, therefore, really focus only on the lexical constituents.

Gramley and Pätzold (1992: 66) conclude that only lexical combinations with a very restricted range, i.e. frozen collocations present a linguistic challenge.

2. Idioms
2.1. Definition

In linguistics, idiom is defined more narrowly as a complex lexical item which is longer than a word form, but shorter than a sentence, and which has a meaning that cannot be derived, inferred from knowledge of its component parts. Meaning is thus the decisive, if not the only criterion for idioms. To contrast collocations and idioms, we can say that collocations consist of two word forms which are at the same time semantic constituents or lexical units, while the word forms in an idiom do not constitute lexical units and do not make an isolable contribution to the meaning of the whole.

An idiom is a “lexical complex which is semantically simplex” (Cruse 1986: 37), or one can say that idioms show unitary meaning. A test for a semantic constituent is that of recurrent semantic contrast (Cruse 1986: 26-29). In the sentence You need not jump down my throat (‘reprimand or contradict a person fiercely’), take need and substitute for it the semantically different, but syntactically identical item may. This changes the meaning of the sentence, but the point is that the same substitution of forms in a completely different sentence will produce a parallel change of meaning, as in they need/may not sit the exam. The same test shows that you is also a semantic constituent, but that throat is not, as in semantically unacceptable *you need not jump down my wind pipe.

Recurrent semantic contrast does not mean that all idioms are equally difficult to decode. Idioms show different degrees of semantic opacity.

2.2. Degrees of opacity

A semantically non-transparent expression may be described as semantically opaque. Transparency as defined by Cruse (1986: 39) is the end – point of a continuum of degrees of opacity. Cruse has located the decisive break in semantic character between “fully transparent” and “to some degree opaque” rather than between “completely opaque” and “not completely opaque”, as this groups together more satisfactorily elements with significantly similar properties. The idea of semi-opaque expressions is already implicit in the notion of semantic indicator; a semi-opaque expression must contain at least one semantic indicator.
There would seem to be two components to the notion of *degree of opacity*. The first is the extent to which constituents of opaque expressions are “full” semantic indicators. The other factor affecting degree of opacity is the discrepancy between the combined contribution of the indicators, whether full or partial, and the overall meaning of the idiom. It is difficult to measure such a discrepancy objectively.

As degree of opacity diminishes, we approach the somewhat indeterminate transitional zone between opacity and transparency. In principle, all opaque sequences are minimal lexical units and therefore should be listed separately in an ideal dictionary (Cruse 1986: 40). Knowledge of the world (encyclopaedic knowledge) plays a part in the degree to which speakers feel idioms to be opaque. Many idioms have originated in metaphors which some speakers recognize, while others remain unaware of their origin. Thus, while *bury the hatchet*, *give somebody the green light* and *gnash one’s teeth* are likely to be known and easily intelligible to many, only few will know that *white elephant* (‘something expensive which is completely useless’) apparently derives from a King of Siam who used to make a present of a white elephant to people he wished to ruin.

Few readers will apply the recurrent semantic contrast test, so there raises the question of what other means there are to recognize that an expression is an idiom. Many idioms have two meanings, a literal and an idiomatic one (*kick the bucket*, *go to the country*, *pull one’s leg*). In such cases only the context can give a clue to which meaning is intended. In other cases, when a literal reading does not make sense in terms of the world as we know it, the likelihood is that we are dealing with an idiom. This applies to *jump down someone's throat*, *fly off the handle*, and *it rains cats and dogs*. The same conclusion should suggest itself when an expression is formed in a way that is contrary to the syntactic rules of contemporary English, as is the definite article in *kick the bucket* or *one in pull a fast one*. The definite article normally has the function of indicating that an item has already been mentioned or is considered unique in the context of the language community, while the pro-form *one* refers to a noun that must precede it. Neither of these conditions is fulfilled in the idioms cited (Gramley and Pätzold 1992: 73).

Finally, idioms can be phonologically irregular in that they have an unpredictable stress pattern. Also, in connected spoken discourse, idioms are often signalled by slight pauses or an audible intake of breath (Strässler 1982: 95). There is also a certain amount of lexical repetition in the environment of idioms, which makes for greater lexical cohesion. Often a state of affairs is described, then the sender refers to it with an idiom, before it is picked up again by a non-idiomatic, literal, lexical item (*idem*: 96).

### 2.3. Idioms and collocations

The term *collocation*, as we have already discussed it, is used to refer to sequences of lexical items which habitually co-occur, but which are nonetheless fully transparent in the sense that each lexical constituent is also a semantic...
constituent. Collocations like fine weather, light drizzle, high winds are easy to distinguish from idioms; nonetheless, they do have a kind of semantic cohesion – the constituent elements are, to varying degrees, mutually selective. The semantic integrity or cohesion of a collocation is more marked if the meaning carried by one (or more) of its constituent elements is highly restricted contextually, and different from its meaning in more neutral contexts (Cruse 1986: 40).

Semantic cohesiveness is even tighter if the meaning of one of the elements of a collocation requires a particular lexical item in its immediate context.

Collocations like foot the bill and curry favour whose constituents do not like to be separated are termed by Cruse as bound collocations. Although they display same of the characteristic properties of idioms, bound collocations are nevertheless, lexically complex (idem: 41).

2.4. Idioms and dead metaphors

There is a type of expression which is frequently included in the category of idiom, but which, according to Cruse (1986) ought to be kept distinct, and that is what is sometimes called frozen or dead metaphors.

If a metaphor is used sufficiently frequently with a particular meaning, it loses its characteristic flavour, its capacity to surprise, and hearers encode the metaphorical meaning as one of the standard senses of the expression. Interpreting it then, no longer requires the activation of the metaphorical strategy, working through the literal meaning, but merely requires the looking up, as it were, of a dictionary entry, in much the same way, presumably, that idioms are interpreted. However, very often the link with the original live metaphor, and hence with the literal meanings of the parts, is not wholly lost. Dead metaphors for which this is true, can be “revived” by substituting for one or more of their constituent parts elements which, in their literal uses, are near-synonyms, or paraphrases. Cruse (1986: 42) considers the following pairs of sentences:

a. They tried to sweeten the pill.
   b. They tried to sugar the medicine.

a. We shall leave no stone unturned in our search for the culprit.
   b. We shall look under every stone in our search for the culprit.

The first sentence in each pair contains a dead metaphor; in the second sentence, the metaphor is revitalised by the substitution of a near-synonym or paraphrase. The same process carried out on true idioms dramatically demonstrates the difference between the two types of expression:

a. John pulled his sister’s leg.
   b. John tugged at his sister’s leg.

a. Tonight we’re going to paint the town red.
   b. Tonight we’re going to colour the city scarlet.

Something similar happens on translation. A literal rendering of an idiom is very rarely capable of serving as even an approximate translation. But literal translation fares rather better with dead metaphors; the results are usually a little odd, but are nonetheless interpretable in the manner of live metaphors. In the
Semantic Transparency And Opacity In Fixed Expressions

following, the (a) sentences are dead metaphors, and the (b) sentences are literal translations (Cruse 1986: 43).

a. Why keep a dog and bark yourself?
b. Pourquoi avoir un chien et aboyas soi-même ?

a. You’re barking up the wrong tree.
b. Ce n’est pas à cet arbre là que vous devez aboyer.

Interestingly, a high proportion of dead metaphors have similar (although not often identical) dead metaphor equivalents:

a. to put the cat among the pigeons.
b. mettre le loup dans la bergerie

Not surprisingly, dead metaphors as a rule present fewer problems to foreign learners of a language than idioms do.

While idioms and dead metaphors must be distinguished, it should also be recognised that they have certain characteristics in common. It is likely that most of idioms began their lives as metaphors; and synchronically, transitional cases, which are idioms for some and metaphors for others, are not uncommon. Dead metaphors have in common with idioms the fact that their constituent elements do not, in the straightforward sense, yield recurrent semantic contrasts. They are not, therefore, semantically transparent. Yet, the effect of synonymous substitution and the continuing relevance of their literal meaning make it unsatisfactory simply to call them “opaque”. Cruse (1986: 44) therefore, describes them as ‘translucent’, but he also notes that translucency is not the same as the semi-opacity. Dead metaphors also have a syntactic rigidity, since the quality of being “dead” is closely related to a particular syntactic form, and with any modification the metaphor springs to life: compare He has one foot in the grave and One of his feet is in the grave.

Because of their non-transparent and syntactic frozenness Cruse (idem: 45) considers dead metaphors to be minimal lexical units.

We can conclude that there are certain degrees of semantic transparency, and it is obvious that, from this point of view, the segmentation of the terminological continuum in phraseology is a very difficult task.

REFERENCES


The paper focuses on the problems of semantic transparency and opacity in fixed expressions, particularly in collocations and idioms. There are degrees of semantic transparency in collocations. Collocations are perceived as being semantically transparent. Yet, this cannot be taken to imply that their meaning is always a compositional function of the meanings of their constituents.

On the other hand, recurrent semantic contrast does not mean that all idioms are equally difficult to decode. Idioms show different degrees of ambiguity, of semantic opacity. Knowledge of the world will play a part in the degree to which speakers feel idioms to be opaque.

Key words: collocations, idioms, semantic transparency and semantic opacity