

## Collocations as one particular type of conventional word combinations Their definition and character

### Abstract

The concept of "collocation" is still somewhat vague, despite attempts over the past few years to narrow it down. There are many kinds of recurrent combinations of words and the present article aims at a further demarcation of the concept within this larger field, partly involving a case study of the words *commit* and *perpetrate*. In addition some attention is paid to collocations across languages: a sample is presented of a Dutch-English collocating dictionary.

Keywords: collocations, word combinations, bilingual dictionary of word combinations

### 1. Introduction

Over the past few years linguistics, and in particular lexicography, has paid an increasing amount of attention to the way in which words habitually group together in clusters that are not considered idioms proper but are yet felt to be frequent and apparently belonging to the set of ready-to-hand units of language comprising more than one word. To varying degrees the English learners' dictionaries have tried to cater for the needs of foreign learners. And, of course, there is the very useful Benson, Benson & Ilson (=BBI) (1997), which is entirely devoted to collocations. Despite this focus on collocations the concept is still somewhat vague and wants a clearer definition or demarcation. The first part of this article will be devoted to an attempt at such a definition (i.e. of lexical collocations). The second part will briefly go into some practical problems arising in this connection in bilingual lexicography (in this case Dutch-English). There is a great need for bilingual collocation dictionaries, which alas still largely remain to be written.<sup>1</sup>

### 2. The concept of "collocation"

Over the past few decades an increasing amount of attention has been lavished on habitually occurring word combinations that are felt to be neither idioms - which are often said to be not fully semantically transparent<sup>2</sup> - nor totally free combinations. I am here referring to combinations like *commit murder*, *follow somebody's example* or *take a decision*. These are ready-made linguistic building blocks larger than words that are used (by the native speaker) as units of form and meaning. These are, in Sinclair's words (1991: 110) a propos of the "idiom principle" "semi-preconstructed phrases that constitute *single choices*, even though they might appear to be analysable into segments" [emphasis mine].<sup>3</sup>

As may be expected, *collocations* - as many linguists would call at least a number of the fixed combinations referred to - are another case of *gradience*: in the same way as not all idioms are equally opaque, not all examples of what are called collocations are equally transparent. In this light it is no wonder that opinions on the demarcation of this concept also vary. Thus, for Hausmann (1985: 118) collocations are typical, specific and characteristic word combinations,

which he calls prefabricated or at least semi-prefabricated (“Fertigprodukte” and “Halbfertigprodukte”), and he adds that obviously not all combinations qualify for the term collocation. Thus, he thinks that a “banal” combination like *ein Buch kaufen / buy a book* is not really a collocation, presumably because it is not ‘typical, specific and characteristic’ enough.

In Benson, Benson, Ilson (1986: 253) we find another interesting discussion of the concept of collocations. They define collocations as “loosely fixed combinations” of the type to *commit murder*, taking up a position between “idioms, on one hand, and free combinations, on the other”. Collocations are “psychologically salient”... “fixed phrases” (idem) differing from idioms in being semantically transparent, and from free combinations in being *frequent* and *not freely variable by means of synonyms*. This latter point (the impossibility of the use of synonyms) seems to imply that there is a certain amount of *arbitrariness* in the choice of collocators: out of a number of potential (“synonymous”) candidates only a subset is normally selected.<sup>4</sup> Thus, the authors assert, the only acceptable synonym, if at all, for *commit* in our example seems to be *perpetrate*, which does not normally occur with *murder* (idem: 253).

Collocations, therefore, are in this light to be regarded as *preconstructed* combinations of words, where the constituent words are *semantically literal* (i.e. *non-metaphorical*).<sup>5</sup> Because they are preconstructed they are conventional, i.e. frequent and psychologically salient. Due to their semantic literalness the meanings of collocations can be arrived at through normal analysis. We would then have the three following possibilities in the case of word combinations:

**FREE COMBINATIONS are:**

not preconstructed, and semantically literal (i.e. the words have retained their conventional literal meanings)

**COLLOCATIONS are:**

preconstructed, and semantically literal (i.e. the words have retained their conventional literal meanings)

**IDIOMS are:**

preconstructed, not semantically literal (i.e. the words, or at least one of them, have NOT retained their conventional literal meanings, or at least cannot be analysed as such)

Here we clearly see the hybrid character of collocations, sharing with idioms their preconstructedness, and hence conventionality, and with free combinations their non-metaphorical character.

A recent and wide-ranging discussion of collocations can be found in Herbst 1996. Herbst studies three approaches to the problem of collocations: the text oriented approach, the statistically oriented approach and the significance oriented approach. His own preference is ultimately for the latter<sup>6</sup>, the significance-oriented approach, which he himself associates with Hausmann. Herbst goes into such questions as whether collocations can be explained semantically, by giving the constituent words such semantic definitions that they naturally fit together (by means of mutual selection preferences) or whether there is a certain arbitrariness in the combinations. After a somewhat confusing discussion he first narrows down the concept of collocation to two rather conflicting types: 1. motivated collocations, e.g. *flock of sheep* etc., where on the basis of the meaning of *flock* words like *sheep* are natural collocates, and 2. what he calls unmotivated collocations, like *commit a crime* or *false teeth*. The problem I have with the latter type is that it is far from clear to what extent, and in which sense, the combinations are unmotivated. Thus,

contrary to what Herbst - referring to Palmer - claims, *commit* does not only combine into perfect combinations with *crime*, *sin*, *suicide* but also according to the *BBI* and *The Guardian* (on CD, 1996) with *burglary*, *theft*, *robbery*, strangely enough considered unacceptable by Herbst. Seen in this light, a definition of *commit* as involving '+ serious crime or something morally wrong'<sup>7</sup> would easily place combinations with *commit* in the first - motivated - category. And the same might hold for many more putative unmotivated collocations.

In his conclusion Herbst narrows down his concept of collocations to this single type (at least if I understand him correctly): the defining criterion is "a certain lack of semantic predictability or transparency". A little further on he speaks of collocations as "idiosyncratic" and "knowledge about the individual word that cannot easily be described in general rules". This view of collocations makes them in my view somewhat similar to idioms, in that the *combination* of the words is not quite transparent or predictable, in the sense that certain combinations happen to exist and others *which might be expected with equal justification* do not. In other words, we are back to the concept of the *arbitrariness* of a combination - or, alternatively, *the arbitrariness of the absence of a combination*. Thus, in Herbst's view, *commit burglary*, which he (erroneously!) considers unacceptable, would be an example of the arbitrary absence of a collocation. It would then perhaps be quite profitable to make *not the combination itself* but the genuine arbitrary absence of other combinations one of the defining criteria. Below, I will examine this crucial question on the basis of a case study (the words *commit* and *perpetrate*, alleged synonyms with arbitrary collocational properties).

It has, at any rate, become clear by now that the concept of collocation means various things to various people, which is not really surprising, since discussions of the concept usually involve criteria that do not logically imply each other. To the extent that the recurring, but logically independent, criteria of either *preconstructedness*, *semantic literalness*, or *arbitrariness* (in the choice of potentially available collocators) receive more emphasis the definitions differ.<sup>8</sup> Since there are usually no clear-cut distinctions in natural language itself it is futile to search for absolutely clear-cut definitions. This should not be interpreted as defeatism, however, for one might - to a large extent in the light of the foregoing - try and define the *prototypical* collocation, which could tentatively be defined as a combination of:

1. two or more lexical units, with meanings also occurring independently elsewhere (in other combinations);
2. which are used *non-metaphorically*,
3. which combination occurs *repeatedly* and *normally* in a language (cf. Carter 1987: 47), as a *conventional building block*,
4. which the language user has available as a whole, to express *conventional established concepts*,
5. whose constituent words are typically in a *grammatical modifier - modified relation* (including that of verb-object),
6. whose constituent words (in spite of point 1!) naturally select each other because the sense definition of the modifier includes the modified (and sometimes vice versa) in a non-banal way (*semantic motivation*)
7. which typically function as *part of a larger group* and not as a complete utterance (sentence) itself.

The above definition leaves numerous conventional groups of words that are not collocations in this view (nor are they necessarily always idioms in the proper sense). Carter (1987: 60) lists the

following “types of fixed expressions” (after Alexander 1984): idioms, which are either unanalysable into separately usable words or in which at least one word is used metaphorically (*spick and span, smell a rat*), proverbs, as complete utterances (which may also frequently be metaphorical), in which the words moreover do not really necessarily select each other in a meaningful way (*A watched pot never boils*), stock phrases, i.a. because the words do not really select each other in a meaningful way (*When all is said and done, unaccustomed as I am to ..*), catchphrases, being complete utterances or because the words do not really select each other in a meaningful way (*That's another fine mess you got me into*), allusions and quotations, idem (*You've never had it so good*), idiomatic similes, being unmotivated (*as sober as a judge*), and discursal expressions, being complete utterances or because the words do not really select each other in a meaningful way (*How do you do, mark my words, we'll now take questions from the floor, Ladies and gentlemen, I thought you'd never ask*).<sup>9</sup>

The above definition is, in addition, very restrictive, since it excludes the hybrid cases in which the modifying word is used in a highly specialised sense (cf. Howarth 1996: 47 on “restricted collocations”), where “specialised” may mean technical, figurative or delexical. Cowie *et al.* (1983: xiii) also speak of restricted collocations (which are sometimes referred to as “semi-idioms”), when one word has “a figurative sense not found outside that limited context”: *jog one's memory, a blind alley, etc.* Judging from the examples, it is always the modifier which is meant here.

Let us now look at some possible candidates for the tag ‘collocation’, as chosen at random from the BBI.

- *feature(s)*: qualifying adjectives include *characteristic, distinctive, special, redeeming, coarse, delicate, fine, regular, soft, optional*. Here the question is how to decide apart from statistical evidence, for why should for example *characteristic* select *feature* in any meaningful way? And why are certain adjectives (e.g. *exceptional, permanent, defining, critical* and also *key (feature)*) excluded in the BBI, to mention just a few.
- *rebuke*: with adjectives *mild, scathing, sharp, stern, stinging* and verbs *administer, deliver, give and draw, receive*. Why is *fierce* not listed, or *severe* or *good-natured*? Also, *rebukes* may be *public* and *pointed* or *swift*, despite the BBI.

The purpose of this exercise is not to criticise the BBI, but just to show that it is very simple to find combinations that intuitively might also seem to qualify. The point is how to decide whether a combination is ‘free’ or a collocation. If the only criterion is a higher-than-expected occurrence rate of the combination, statistical evidence would be the only way to settle this problem, which might yield numerous ‘banal’ collocations.<sup>10</sup> As remarked above, such evidence is only heuristic, for in my view a high rate of occurrence is due to, and not the cause of, psychological salience. In other words, because the combination is a building block it occurs more often than statistically expected, and not the other way round. If this were not so, nothing would be able to explain the frequency of occurrence in the first place.

The question then becomes “why is this a building block, and why exactly in this form”? The answer to the first half of this question could be: because there is a need for it for language *external* (the world around us) reasons: these combinations are needed because we need them to discuss the topics we wish to discuss. In other words: between the level of the word (i.e. the conventional association of form and meaning) and the level of the sentence (i.e. a non-conventional association of form and meaning)<sup>11</sup> there are quite obviously all kinds of more or less conventional associations of form and meaning, with various degrees of “frozenness” and

cohesion.<sup>12</sup> That is, we use all these building blocks not only because they are there, but for the deeper reason that we use them because they conventionally express conventional *ideas*, *concepts* or whatever we like to call them. The answer to the second might be: they are simply the available ways to speak about the desired topics. If there are other ways, these are in principle also used. However, the more settled, conventional and circumscribed the concepts are, the more settled, conventional and circumscribed the expressions generally become. Thus, *the white house* and *the White House*, *the red book* and *the Red Book* (of Chairman Mao) differ in that the capitalised combinations express well-defined and well-known conventional concepts, causing what are in principle free combinations to acquire additional meanings so that their constituent words no longer suffice to independently and fully “motivate” the expressions. The *White House* with its accreted meanings of ‘seat of government’ and ‘residence of the US president’ is therefore not a (prototypical) collocation under my definition.

The existence of perfect possibilities of combination that are nevertheless not used (the *arbitrariness* argument) is well worth discussing in some more detail. One such example might be some combinations with either *dangerous* or *hazardous*. If waste is dangerous, it is almost always called *hazardous waste*, and only rarely *dangerous waste*. Other common combinations like *hazardous chemicals*, *journey*, *operation*, *undertaking* may be common with *hazardous*, but *dangerous* is certainly found as well here. In my view, *hazardous waste* is, due to certain aspects of our modern world, the more clearly conventional and therefore the more clearly defined unitary concept, which will probably lead to “blocking”, a mechanism known in morphology: if we already have e.g. *transparency* as an established word for a concept, there is no need to also have *transparentness*, which therefore occurs but rarely (though it is formed with a fully productive suffix). I suspect that in this case we have something very similar: if there is a clearly defined concept there is no need to have two rather similar expressions for it.

Let us examine another pair of putative candidates for the label “synonymous”, the words *commit* and *perpetrate*, already mentioned earlier as being synonyms though with different collocation patterns. This pair may nicely illustrate the problems we have been dealing with. The COD<sup>13</sup> defines *commit* - in the sense that concerns us here - as ‘perpetrate, do (esp. a crime, sin or blunder)’ and *perpetrate* as ‘commit or perform (a crime, blunder or anything outrageous)’. These definitions would make a good case for considering them synonyms.

Let us now study the results of the evidence I found in *The Guardian/The Observer* (volume 1996, with some additions from 1995) on CD. Here we see that both *commit* and *perpetrate* can combine with the following nouns:

*abuse(s)*, *child abuse*, *act* (of violence, banality etc.), *actions*, *assault(s)*, *atrocities*, *breaches* (of article 14 etc.), *conspiracy*, *crime*, *cruelities*, *damage*, *deception(s)*, *deeds* (murderous, etc ..), *evil*, *excesses*, *follies*, *forgery*, *fraud*, *genocide*, *heist*, *horror(s)*, *indiscretion*, *infanticide*, *injustice*, *massacre*, *murder* (mass m.), *offence* (sexual ..), *outrage*, *rapes*, *theft*, *treason*, *violation(s)*, *violence*, *wrong(s)*.

With *commit* only I found (those nouns I consider (legal) terms for ‘official crimes or wrongdoings’ have been capitalised):

*ADULTERY*, *ARSON*, *attack*, *blasphemy* (*Webster*), *blunder*, *BURGLARY*, *contamination* (*product* ..), *corruption*, *customs* (*abominable* ..), *error*, *ESPIONAGE*, *ethnic cleansing*, *EXTORTION*, *faux pas*, *foul*, *FRAUD(S)*, *hara-kiri*, *heresy*, *HOMICIDE*,

*hypocrisy, imprisonment (false ..), INFANTICIDE, injury (personal ..), intercourse (with a girl aged under 13), killings, mayhem, misdemeanour, obstruction, PAEDOPHILIA, PERJURY, practice (unpleasant sexual ..s), push shot (!), ROBBERY, sex (oral ..), sin, sloth, gluttony and luxury, solecism, SUICIDE, terrorism, wounding (serious ..)*

and with *perpetrate* only:

*activities, affray, onanism, illegitimacy (with perpetrator), apartheid (with perpetrator), attack(s), attempt, barbarity, begging (aggressive ..), bomb (with perpetrator), bullying, butcheries, campaign (with perpetrator), carnage (with perpetrator), concealment, crossing-over, destruction, dirty tricks, discrimination, disorder (with perpetrator), double standards, dropping litter, error (with perpetrator), falsehood, fictions (great ..), graffiti, harassment (with perpetrator), harm, hit on Citibank, hoax, holocaust, homicide (with perpetrator), humiliation, incarceration (with perpetrator), incident (with perpetrator), insider dealing (with perpetrator), interruptions, joke, journalism (gutter ..), lie, malfeasance (financial ..), miscarriage of justice, miss of the year, monstrosity, myth, mythology, neglect, pogrom, policies of divide and rule, racket, riots (with perpetrator), rip-off, scam, sculptures (with perpetrator), sex abuse, sex crimes (with perpetrator), sex scandal, sham, skipping (mass ..), slaughters, sleaze (with perpetrator), stabbing (brutal ..) (with perpetrator), stupidities, terrorism (with perpetrator), torture, treatment (inhumane ..), trick, view of New York, waste of life, wordplay (tortuous ..).<sup>14</sup>*

We can learn the following from this. The greater “range” of *perpetrate* indicates that its meaning is more general than that of *commit*.<sup>15</sup> Secondly, a comparison shows that *commit*, much more than *perpetrate*, combines with nouns with formal and/or legal import, words referring to what could be called “official” crimes or wrongdoings (I have highlighted these words by capitalising them under *commit*). It will be seen that *perpetrate* shares some of these “official” crimes with *commit* (such as *abuse* - in combinations like *child abuse* -, *arson*, *crime* itself, *genocide*, *murder*, *offence*, *theft*), but since there are very few “official” crimes with which *perpetrate* combines exclusively, it is obviously *commit* which is the more officially sanctioned collocater in as it were multi-word lexical units referring to “official” crimes or wrongdoings, in set phrases like *commit suicide*, *murder* etc. All this means that far from being synonymous, there are good reasons for believing that the two words differ significantly, in spite of a certain extent of semantic overlap. This is, for *perpetrate*, especially clear in cases like *graffiti*, *gutter journalism*, *myth*, *mythology*, *mass skipping*<sup>16</sup>, *view of New York*, where paraphrases like ‘perform’, ‘do’ become rather difficult. Rather, the meaning seems to be ‘to have on one’s conscience (the existence of)’. Yet, even this list shows that *commit* can also be used less officially, cf. e.g. the rather fanciful combinations with *sloth*, *gluttony* and *luxury*. The list of nouns combining with *perpetrate* only is, however, a good deal more informal.

In short, despite the case that has been made for the arbitrariness argument (i.e. a certain word does not collocate though a “synonym” does), the specific evidence presented here urges us to be very cautious. More evidence in the form of case studies like this is needed, though it will often lead to circularity: if word x does not collocate and word y does, word x will be taken to be no synonym, and if it is no synonym its behaviour will not be arbitrary collocation-wise.

I am therefore inclined to believe that what could still be called “collocations” and “free combinations” only differ in that the former are familiar building blocks, which to some extent stand for more or less fixed “concepts” which we need in order to speak about regular features

of the world around us, as in the case of *hazardous waste* or *commit suicide*. Such combinations may thus be felt to be rather like multi-word lexical units. "Banal" combinations like *buy a book, etc* may be relatively frequent, but need not be psychologically salient ready-made building blocks (due to their low information content as combinations they do not express real concepts). Thus, many collocations differ from "free combinations" mainly in that they are available as "given" and "prefabricated" combinations of free words expressing "given" and "prefabricated" concepts. The absence of expected combinations need not always be *arbitrary*, but may after a careful analysis well be due to meaning differences.<sup>17</sup> If, after a serious examination of the facts, no reason can be found for the absence, or even perceived incorrectness, of a combination which would make perfect sense, we would have an example of "blocking": a well-established concept being expressed by a well-established combination.

This still leaves us with a concept that is fuzzy around the edges. But perhaps this is exactly as it should be, given the nature of language: some combinations will be more fixed than others because some concepts are better established than others. There can be no doubt that *to perpetrate a view of New York* is a free combination, whereas *to perpetrate a crime* had better be called a collocation. As a heuristic method, statistics will be useful to solve this problem of where collocations stop and free combinations begin. But statistics, too, have to be interpreted and even here there will remain grey areas. Language is usually a matter of more or less, and giving names to relatively clear phenomena does not make less clear phenomena suddenly much clearer. Language offers constraints (i.e. clear collocations) but also offers us freedom: the freedom to leave constraints and strike out on our own with the constraints as our compass. Thus, a knowledge of constraints makes possible *to perpetrate a view of New York* (as opposed to the much more unlikely *to commit a view of New York*): a knowledge of the usual collocational range of *perpetuate* - and hence indeed of its meaning - teaches us that this *view of New York* can never be a good thing.

### 3. Word combinations and bilingual lexicography

Two years ago I wrote an article on collocations (Van der Meer 1997), for which I did research on all possible Dutch combinations with the word *angst* 'fear' and their English equivalents. It was a very enlightening experience to discover to what extent the two languages, seen from the point of view of Dutch as the source language, did, but also did *not*, correspond. What quickly became evident was the degree to which Dutch combinations could often not be translated literally, at least not into idiomatic English. One of the reasons for this is no doubt the frequently only partial semantic overlap of translation "equivalents". An ideal bilingual dictionary should, therefore, not only offer target-language combinations that are totally unpredictable from the Dutch viewpoint because they do not correspond to the Dutch pattern at all, but also those that - on the basis of partial equivalence of one of the words involved - might with some reason be expected also to correspond, but in fact do not. Take for example the Dutch adjective *scherp*, often corresponding to English *sharp*. Take, for instance, *scherpe concurrentie* (lit. 'sharp competition'), which had better be rendered as *tough* or *keen competition* and not *\*sharp competition*.

I will here print part of a possible lemma from a Dutch-English combinatory dictionary as it might - and should! - be compiled. The ordering is simply alphabetic. In a number of cases there may, of course, be differences of opinion as to the proper "equivalents".

ANGST (*zie ook vrees*)

♦ fear (of), anxiety (about, over), <sterker> dread, terror (of), ♦ (*psychologische* ≈) anguish (at), ♦ (≈ *voor wat gaat gebeuren, het onbekende*) dread (of), ♦ (*neurotische* ≈) angst; ▣ *in verbindingen* (*zie ook vrees*) (= *fear tenzij anders aangegeven*) (*iem.*) a. >aanjagen frighten; <sterker> terrify; put the fear of God into sb, to inspire fear (*of*: terror), put (*of*: strike) fear in the hearts of (*bijv* youngsters); >aanwakkeren add to; >achtervolgen haunt; >afleggen, shrug off; >afnemen: diminish, subside; (*van je*) >afzetten shrug off; >algemene: common, public, prevailing (*ook de a. is algemeen dat* fears are rife that); >beheersen control; >bekruipen creep over (s.b.); >bestaan (*bij*) (people) suffer fear, have a fear of; *met a. en* >beven (*iets tegemoet zien*) (*bijv* view (*of*: await) sth.) with terror in one's heart, .. with fear and trembling, (*bijv* anticipate sth.) with (*of*: in) fear and trembling; >bevorderen promote; >bewaarheid worden be confirmed, realised; >bezweren allay; >bezwijken *van bijv* I almost died (*of*: passed out) with fear; >bij bij (*de mensen*): fear among (*bijv* the people); >blijven bestaan persist; *de* >boventoon voeren (pre)dominate; >diepe profound, deep, grave; >diepgewortelde deep-seated; >dodelijke deadly, mortal; >doen afnemen reduce, diminish, relieve; >doen herleven (*of*: opleven) renew, revive; >doen verdwijnen dissolve; >doen toenemen add to, increase, heighten; >door (*bijv* paralysed) with (*of*: by) fear, shaken with pain and fear; >doorstaan suffer fears, be terrified, be anxious; >duizend angsten, *zie uitstaan*; >enorme: tremendous; >houden (*de a. erin* ≈) perpetuate; >gegronde reasonable, well-founded, well-grounded; (*zonder ook maar de*) >geringste: without the slightest fear; *een* >golf *van* a wave of fear; >groeien grow, increase; >groeie *ook* mounting; >groot zijn: *bijv* fears are running high; >grote big, great, grave; *iem. om het* >hart slaan: to be gripped by; *met a. in het* >hart with fear in his heart; >hebben (*voor*) be afraid, scared (of), <*schr.*> be (*of*: stand in) dread (*of*: fear) of; >hebben *ook* experience, have (got) a fear (*bijv* of failure), be in fear of; >heersen rule (*bijv* >er heerst a. dat there are widespread fears that); >hevige (*of*: heftige) strong, tremendous; >herleven: revive; >in a. verkeren, zijn, zitten e.d. be afraid (*of*: scared), be (go, live) in fear (and trembling), in terror, <*informeel*> be in a fright (*of*: panic), be under fear; etc. etc.

## 4. Notes

- <sup>1</sup> From Fontenelle 1997 I learned of the existence of Ilgenfritz *et al.* 1989, as one of the few exceptions.
- <sup>2</sup> I should point out here that this cannot be taken at its face value, for - as has been amply demonstrated in Gibbs 1990 and Verstraten 1992 - what we feel are idioms are quite often not so much unanalysable as non-literal. Quite frequently idiomatic expressions are easy to analyse as metaphors, especially when they are instantiations of underlying conceptual metaphors like ANGER IS HEAT or ARGUMENT IS WAR (also cf. the groundbreaking Lakoff & Johnson 1980). So the point seems rather to be that collocations are non-metaphorical.
- <sup>3</sup> Note Bolinger's wise words (Bolinger 1975: 105) to the effect that 'the brain stores both the parts and the wholes, and we retrieve them when we need them'.
- <sup>4</sup> The word 'arbitrariness' is explicitly used by Mel'chuk & Wanner 1994: 325. They also hint at the unpredictability or lack of motivation of the collocations, though they concede that there may be 'partial motivation' (idem). Yet, idiosyncracies abound in collocations (idem: p. 326), 'so that many of them have to be simply listed'. Also cf. Benson's own definition in 1989, p. 3: collocations are 'arbitrary, recurrent word combinations'. Benson also points out that the arbitrariness of 'collocations' is especially clear when languages are compared in this respect, something every learner will confirm.
- <sup>5</sup> I write 'literal', and not 'transparent' as the authors above said. The two need not necessarily mean the same, but for the time being I will try and see how far 'literal' will get me.
- <sup>6</sup> The first boils down to the rather unrevealing insight that one may expect cohesion in texts about a certain well-defined subject: in a text about coastal walking one may indeed expect words like *coast, path, sea*,



*climb* etc. The second approach is mainly associated with John Sinclair's work, who distinguishes between casual and significant collocations. The latter occur more frequently than might be expected statistically.

- 7 This would also make *commit suicide* a typical collocation, cf. e.g. CIDE's definition 'to do (something illegal or considered wrong)'. CIDE gives examples with *crime, sin, murder, offence*. LDOCE3 gives almost exactly the same definition, and the collocates mentioned are: *crime, murder, rape, adultery* etc. [!] and also *suicide*. All this clearly supports placing *commit* in the first, motivated, category.
- 8 The general feeling that collocations represent the normal and natural ways of combining words in a language may, according to Carter, also be explained by using the concept of 'range' (Carter 1987: 52, who borrows this idea from McIntosh 1966: 189ff.). Just as in grammar (syntax) there are 'patterns' determining the way (exponents of) grammatical classes naturally combine, there are 'ranges' for individual words, explaining how they naturally combine. Thus, there is a 'range [...], which is represented by the fairly limited inventory of nouns which may without any question be qualified by the word *molten*. The set of alternative available possibilities which this inventory consists of is just as much a part of the form of the language as is a grammatical system, and a full account of this set goes a long way towards constituting the meaning of *molten*' (McIntosh 1966: 189). Carter (1987: 52-3) mentions the four words *putrid, rotten, rancid, addled*, which though being 'all virtually synonymous' collocate with quite different nouns. If, however, as McIntosh argues a propos of *molten*, the meaning of a word is also defined by its range, it would in my view be an exaggeration to speak of 'virtually synonymous' here. Again, it is the vagueness of the word *synonymous* which helps to contribute to the definitional problems of the concept of collocations itself. Consequently, the concept of range is not very helpful if the concept of synonymy is used in such a debatable manner.
- 9 Respectively subcategorised as 'social formulae/clichés', 'connectives; structuring devices', 'conversational gambits', 'stylistic formulae' and 'stereotypes'.
- 10 Which still does not solve the problem how much the expected occurrence rate must be exceeded for the combination to count as a 'collocation'.
- 11 With the obvious exception of sentence-sized expressions and proverbs.
- 12 If we, in Bolinger's words, were 'to analyze things down to the smallest bits and then put them together again with grammatical rules', as a lot of not only American linguists would have us believe, speaking would involve an intolerable amount of calculation (Bolinger 1975: 102). Thank God for conventional language!
- 13 *The Concise Oxford Dictionary*, Ninth Ed., Oxford: The Oxford University Press, 1995.
- 14 I must here of course stress that not all nouns included in the above three lists should be considered 'collocates', for not all of them need to be considered ready-made familiar building blocks.
- 15 Note that some of these words were only found with *perpetrator*, which is more freely usable due to the fact that \**committer* is not a possible word.
- 16 Meaning apparently 'mass queue jumping'.
- 17 The risk of circularity should be avoided here by making the analysis as complete as possible, cf. the *commit-perpetrate* case.

## 5. References

- Alexander, R. J. (1984). "Fixed Expressions in English: Reference books and the teacher"; *English Language Teaching Journal*, 38: 127-34.
- Benson, Morton (1989). "The structure of the collocational dictionary"; *International Journal of Lexicography* 2: 1-14.

- Benson, Morton, Evelyn Benson & Robert Ilson (1986). *Lexicographic Description of English*; Amsterdam.
- Benson, Morton, Evelyn Benson & Robert Ilson (1998). *The BBI Dictionary of English Word Combinations*; Amsterdam/Philadelphia: John Benjamins.
- Bolinger, Dwight (1975). *Aspects of Language* (second ed.); New York etc.: Harcourt Brace Jovanovitch.
- Carter, Ronald (1987). *Vocabulary. Applied Linguistic Perspectives*; London: Allen & Unwin.
- Cowie, A P, R Mackin, I R MaCaig (1983). *Oxford Dictionary of Current Idiomatic English. Volume 2: Phrase, Clause & Sentence Idioms*; Oxford: Oxford University Press.
- Fontenelle, Thierry (1997). *Turning a Bilingual Dictionary into a Lexical-Semantic Database*; Tübingen: Max Niemeyer Verlag (Lexicographica Series Major 79).
- Gibbs, Raymond W (1990). "Psycholinguistic studies on the conceptual basis of idiomaticity"; *Cognitive Linguistics* 1-4: 417-51.
- Hausmann, F.J. (1985). "Kollokationen im deutschen Wörterbuch. Ein Beitrag zur Theorie des Lexikographischen Beispiels"; in Bergenholtz/Mugdan (eds.), *Lexikographie und Grammatik* 118-129.
- Herbst, Thomas (1996). "What Are Collocations: Sandy Beaches or False Teeth?"; *English studies*, Vol. 77 (1996), no. 4, pag. 379-393.
- Howarth, Peter Andrew (1996). *Phraseology in English Academic Writing. Some Implications for Language Learning and Dictionary Making*; Tübingen: Max Niemeyer Verlag (Lexicographica Series Major 75)
- Ilgenfritz, P., N. Stephan-Gabinel & G. Schneider (1989). *Langenscheidt's Kontextwörterbuch Französisch-Deutsch*; Berlin-München: Langenscheidt
- Lakoff, George & Mark Johnson (1980). *Metaphors we live by*; Chicago: Chicago University Press.
- McIntosh, Angus (1966). "Patterns and ranges"; in McIntosh, Angus, Halliday, M.A.K. (1966). *Patterns of language : papers in general, descriptive and applied linguistics*; London: Longmans.
- Meer, Geart van der (1998). "Het actieve vertaalwoordenboek en collocaties" ; *Trefwoord* 12: 134-142.
- Mel'čuk, Igor & Leo Wanner (1994). "Towards an Efficient Representation of Restricted Lexical Cooccurrence". In *Euralex 1994 Proceedings. Papers submitted to the 6th EUR-ALEX International Congress on Lexicography in Amsterdam, The Netherlands*, pp. 325-338.
- Sinclair, John (1991). *Corpus, Concordance, Collocation*; Oxford.
- Verstraten, L P (1992). *Vaste verbindingen*; Utrecht: LED