THE CASE FOR VERB-ADJECTIVE COLLOCATIONS: CORPUS-BASED ANALYSIS AND LEXICOGRAPHICAL TREATMENT[1]

Moisés Almela
Universidad de Murcia

Abstract: *This article explores a type of co-occurrence pattern which cannot be adequately described by existing models of collocation, and for which combinatory dictionaries have yet failed to provide sufficient information. The phenomenon of "oblique inter-collocation", as I propose to call it, is characterised by a concatenation of syntagmatic preferences which partially contravenes the habitual grammatical order of semantic selection. In particular, I will examine some of the effects which the verb cause exerts on the distribution of attributive adjectives in the context of specific noun classes. The procedure for detecting and describing patterns of oblique inter-collocation is illustrated by means of SketchEngine corpus query tools. Based on the data extracted from a large-scale corpus, this paper carries out a critical analysis of the micro-structure in Oxford Collocations Dictionary.*

Key words: *Collocation, inter-collocation, lexical constellations, combinatory lexicography.*

1. INTRODUCTION

Despite the ambiguity with which the term *collocation* has been used in the literature,[2] it is possible to point out a number of features which many approaches consider to be characteristic of this phenomenon. Apart from their syntagmatic stability and semantic transparency, the features commonly attributed to collocation include a binary composition and a specific syntactic arrangement. On this view, lexical collocations always consist of two conceptual elements of a certain syntactic type (Martin, 2008: 56). With the exception of British Contextualism (neo-Firthian linguistics), most traditions of collocational research are generally agreed on these terms.

The two features mentioned above are reflected in the design of grammatical typologies of collocation (Hausmann, 1998; Corpas Pastor, 1996, 1998). In most classifications the syntactic realisations of lexical collocations[3] are reduced to five basic types: verb + noun (e.g. *reach an agreement*), adverb + verb (*drink greedily*), adjective + noun[4] (*heavy smoker*), adverb + adjective (*blissfully happy*), and quantifier + *of* + noun (*flock of birds*). All five types have a fundamental characteristic in common, to wit: they involve the combination of a predicative lexeme with one of its arguments (Bosque, 2001, 2004).

[1] This paper is based on research generously funded by a grant from *Fundación Séneca, Agencia de Ciencia y Tecnología de la Región de Murcia* (Ref. 08594/ PHCS/08).

[2] For an overview of different definitions and uses of the concept of collocation, the reader is referred to the following authors: Corpas Pastor (2001), Almela (2006: 118ff).

[3] The reason why I insist on adding the attribute *lexical* here is that some lexicographical works have applied the term *collocation* both to lexical and to grammatical patterns. For instance, in *The BBI Dictionary of English Word Combinations,* a grammatical collocation consists of combinations of lexical items and grammatical tags or function words (e.g. *adjective + to*-infinitive, *noun + preposition*). The phenomenon of grammatical collocation does not lie within the remit of this article.

[4] In some languages like English, it is possible to recognise combinations of two nouns (e.g. *data source*) as a variant of this category of collocation. Hence, the third syntactic type can be represented as *modifier + noun*, rather than *adjective + noun*. That is, the morphological realisation of the modifier can be left unspecified.

In some traditions, there are specific terms for each of the two components of a collocation. Hausmann (1979) coined the terms *base* and *collocator*, which denote respectively the argument and the predicate. The same terminology has been gradually adopted by the proponents of Explanatory and Combinatory Lexicology (ECL), the lexical component of Meaning-Text Theory. Originally, the terms employed in ECL were *key-word* and *value*. However, as Mel'čuk et al. (1995: 126) conceded, the key-word/value distinction is comparable to that between base and collocator, and with time, this second pair of terms (base/collocator) has become usual in the ECL literature.

In this article, I will undertake a critical revision of the conventional syntactic typology of collocation. The goal is to determine whether the received models should be refined or extended to include co-occurrence patterns which share essential characteristics with collocations, but which do not fit well into the established syntactic types.

The paper is formally structured as follows. First, in the next section, I will briefly describe the reasons which in previous research have motivated the reduction of collocation to predicate-argument structure. Then, in section 3, I will analyse in-depth the characteristics of a complex pattern (*caused by* + adjective + noun) and evaluate the way it is accounted for in the most recent collocation dictionary of English, namely, the *Oxford Collocations Dictionary* (henceforth: OCD). A comparison with other combinatory dictionaries has not been included due to limitations of space. Finally, the conclusions from this research are summarised in section 4.

2. COLLOCATION AND SEMANTIC SELECTION

The main evidence advanced in support of a restricted syntactic typology of collocations is the *directionality* of semantic selection. In general linguistics, it is a well-known fact that distributional classes of arguments are semantically more homogeneous than distributional classes of predicates. As a rule, it is easier to predict the semantic class of an argument given a predicate than it is to predict the semantic class of a predicate given one of its arguments. This is generally interpreted as an indication that the semantic class of an argument is restricted by the predicate, but not vice versa (Bosque, 2001, 2004; Cruse, 2004: 223). To put it another way: predicates function as selectors, and arguments as selectees.

The correlation of syntactic structure and semantic selection provides us with a useful framework for an efficient description of collocation, because it allows us to generate multiple collocational expressions from a single combination rule. For example, the expressions sit *an exam, sit a final, sit a practical, sit an A level*, etc., can be derived from a common pattern of semantic valency. The verb sit, when used transitively, exhibits a preference for object nouns which denote the 'process of finding out how much someone knows' (the rule is, of course, probabilistic rather than deterministic, as is the case of all rules concerning the collocational behaviour of words).

By contrast, no rule can be formulated to predict with relative accuracy the semantic class of verbs that collocate with *exam*. Verbal expressions such as *prepare for, sit, pass, fail, administer, mark*, or set, among others, cannot be subsumed under a general semantic heading, apart from the reference to exam itself. They form a distributional class, but not a semantic class. This illustrates a general fact. The collocators of a base must be described individually, while the bases of a collocator can be described systematically (Bosque, 2001, 2004). This is the principle which informs the structure of REDES, the first major dictionary of Spanish word combinations.

Some authors have made the case for a distinction between *semantic* selection and lexical selection arguing that they have opposite types of directionality (Alonso Ramos, 2007; Barrios,

2009, 2010). The notion of *lexical* selection, as opposed to semantic selection, is based on the observation that collocators (predicates) with potentially equivalent meanings have a different syntagmatic distribution, because they are attracted to the contexts of different collocational bases (arguments). For instance, *commence* and *break* out collocate with proceedings and war, respectively, but **proceedings break out* and **war commences are lexically* odd expressions. Nevertheless, it should be pointed out that the distinction between lexical selection and semantic selection is controversial–for an appraisal, cf. Bosque (2001, 2004) and Apresjan (2009).

For the sake of focus, I will not engage here in the debate about the relationship of lexical selection and semantic selection. I will limit my discussion to the topic of semantic selection, with a special emphasis on what it implies for the design of syntactic typologies of collocation. In particular, the problem I address in this section is the observation that semantic selection can operate in grammatical contexts which do not represent typical instances of predicate-argument structure.

A case in point is the collocation of the verb cause with adjectives expressing a negative evaluation, e.g. faulty, defective, abnormal, deficient, improper, etc. Table 1 contains a list with the most significant adjectival collocates found in a +1 window with respect to the expression caused by. The statistical analysis is based on data from a large English web-derived corpus, named ukWaC v1.0 old (size: 1,526,599,198 words). The corpus is accessible through SketchEngine query system. The measure of lexical association I have applied is logDice. The collocates are arranged in order of statistical significance (from top to bottom, first, and then from left to right), and the scores are given in brackets. For a description of the formula and an explanation of the advantages over other association measures, the reader is referred to Rychlý (2008).

faulty (7.608)	*prolonged (5.855)*	*heavy (5.104)*
excessive (7.550)	*accidental (5.791)*	*secondhand (5.097)*
defective (6.507)	*poor (5.655)*	*inaccurate (5.095)*
inadequate (6.497)	*viral (5.628)*	*chronic (5.066)*
incorrect (6.391)	*bacterial (5.617)*	*airborne (5.022)*
abnormal (6.266)	*passive (5.567)*	*contaminated (4.984)*
improper (6.244)	*anti-personnel (5.370)*	*repetitive (4.978)*
insufficient (6.183)	*parasitic (5.289)*	*man-made (4.918)*
inappropriate (6.034)	*infectious (5.153)*	*toxic (4.881)*
careless (5.961)	*torrential (5.118)*	*protozoan (4.879)*

Table 1. Top adjectival collocates of caused by (+1).

The data shown in Table 1 point to a close conceptual similarity among adjectives which occur in agentive *by*-phrases after the verb *cause*. The most significant adjectives found in this context form a broad semantic class characterised by negative connotations. More precisely, it is possible to discern two subclasses: first, adjectives of disapproval, that is to say, adjectives which convey a negative evaluation of the (extra-linguistic) referent denoted by the noun they modify (e.g. *faulty, excessive, defective, inadequate, incorrect, improper, inappropriate*, etc.); and second, adjectives which describe specific properties of nouns relating to unfavourable situations or events (for instance, *viral, bacterial, parasitic*, and *infectious* allude to diseases).

This information is consistent with findings from previous corpus-based research suggesting that *cause* is associated with lexical contexts which denote unfavourable situations or events

(Stubbs, 1995). To use the habitual terminology in corpus linguistics, we can say that *cause* has a *negative semantic prosody*. What I find interesting to underline in this section is the variety of grammatical combinations in which the negative prosody of *cause* is manifested. This prosody is expressed in the combination of *cause* not only with nouns (e.g. *stress, accident, earthquake, defect,* etc.) but also with other word classes, for instance with the adjectives in Table 1. This can be problematic for mainstream models of collocation, because the combination of a transitive verb and an attributive adjective does not represent a typical example of predicate-argument structure.

Finally, it is essential to remember that, in analysing verb-adjective collocations, factors such as the subcategorisation of the verb and its complementation pattern play a decisive role. Luzón and Campoy (2000) carried out an interesting research into the collocations of adjectives with *linking verbs of transition* (e.g. *go berserk*, *become clear, come true*). Adjective phrases are characteristic arguments of link verbs, and in this sense, the collocational patterns described by Luzón and Campoy (2000) are not at variance with the established syntactic typologies of collocation. However, the same does not hold for the kind of verb-adjective collocations I have described in this section (e.g. *caused by faulty + NN, caused by defective + NN,* etc.). Here, the adjective is in attributive rather than in predicative position, and the verb (*cause*) is not a copula.

Admittedly, the agentive *by*-phrases in which these adjectives occur contain a noun phrase which represents an argument of the *cause*. Owing to the diathesis of the verb, the collocations found in this context are equivalent to those found in the subject slot. Examples (1) and (2) below are different morphosyntactic realisations of the same deep argument structure. However, since the adjective is not the head constituent of any of these argument phrases, the syntactic form of these collocations must be dealt with as a special case.

(1)*...that a disease is caused by a faulty gene...*
(2)*...have now identified the faulty gene which causes the disease...*

One way of dealing with this complexity is to explore the interactions of different collocations. This requires us to take a step from an intra-collocational to an inter-collocational perspective– that is, from the analysis of dependency relations between different components of a collocation to the analysis of dependency relations between different collocations. The proposal is developed in the next section.

3. FROM COLLOCATION TO INTER-COLLOCATION

As I have anticipated in the previous section, the base of a collocation does not necessarily coincide with the head of the syntactic constituent which instantiates the argument. Therefore, one of the difficulties with which the conventional syntactic typologies of collocation must be faced up is the mismatch between the binary structure of collocation and the complexity of the possible morphosyntactic realisations of an argument slot. Let me illustrate the problem by referring to the following examples (concordances extracted from the corpus *ukWaC v1.0 old*, at SketchEngine).

(3a) *…the law of compensation for psychiatric injury caused by occupational stress*
(3b)*...if they are caused by a mental disorder.*

(4a)*...and protects against faults caused by faulty workmanship*
(4b)*...and others in the supply chain, for damage caused by defective goods*

In (3) and (4) we observe different combinations of the verb *cause* (in passive constructions) with agentive *by*-phrases. In analysing this lexico-grammatical pattern (*caused by + ADJ + NN*), the conventional models allow us to recognise a combination of two different word pairs which fit into the syntactic types expected from a collocation: first, the combination of a verb (*cause*) and a noun (*stress, disorder, workmanship, goods*), and second, the combination of an adjective and a noun (*occupational stress, mental disorder, faulty workmanship, defective goods*). At this point there is no gap between collocational associations and phrase structure. The collocation of a noun with an adjective can be embedded into the combination of the same noun with a verb: *[caused [by [[occupational] stress]]]*.

Yet, this analysis is complicated by the observation that only one of the two verb-noun combinations reflects an underlying pattern of semantic selection. The collocation with *stress* is in accord with the negative semantic prosody of *cause* (see Section 2), but the collocation with nouns such as *workmanship* or *goods* is not. There is thus an important difference between example (3) and (4). It is only in example (3) that the selection of the noun is semantically controlled by the verb. However, this does not mean that *caused by faulty workmanship* and *caused by defective goods* are atypical combinations, or that they are not representative of the usage of *cause*. On the contrary, *faulty* and *defective* are highly representative of the negative semantic prosody of *cause* (see Table 1), and in turn, *workmanship* and *goods* are highly representative of the lexical contexts of *faulty* and *defective* (see Section 3.2). The difference between (3) and (4) resides mainly in the grammatical category of the unit which represents the target of semantic selection. In example (3), that unit is also the head constituent of the argument phrase, but the same does not apply to example (4).

The upshot is that the apparent incongruity between verb-adjective collocation and argument structure can be resolved by a careful delimitation of descriptive levels. A syntactic typology of collocation should reflect the difference between the functional description of arguments, on the hand, and the formal-structural description of their realisations at the morphosyntactic level, on the other. Without this distinction, it would be difficult to explain, for example, why in a collocational pattern of the form *verb + NP/PP*, the semantic selection made by the verb can be expressed in constituents other than the head noun.

In the next subsections (3.1 and 3.2) I will proceed to a more detailed examination of the differences observed between (3) and (4). These examples, I will argue, represent different types of inter-collocational patterning. The term *inter-collocation* was introduced by Stubbs (2002: 205) and has later been developed in the framework of the Lexical Constellation Model (Almela, (forthcoming); Almela et al., (forthcoming)). An inter-collocation can be described as a chain of successive collocations with shared components. For instance, the expression *caused by mental disorder* is a function of two collocations that exist independently of one another: *cause + disorder* and *mental + disorder*. The expression *caused by faulty workmanship* is also an example of inter-collocation, but in this case the shared element is the adjective, not the noun. The mechanism of inter-collocation varies substantially from one case to the other, because the collocation of the verb and the adjective seems to be motivated by a compensation for the lack of a strong cohesive relation between the verb and the noun. This is the basic idea developed below.

3.1 Analysing direct inter-collocation

Observe the *word sketches* in Tables 2, 3 and 4 (the corpus is the same one that was used for generating Table 1). A word sketch is an ordered set of significant lexical co-occurrences of a word in specific grammatical relations. In recent versions of SketchEngine, the measures of sta-

tistical salience used in word sketches are based on logDice scores. In Table 1 it was not possible to use the same tool because the system does not generate the necessary word sketches. It does not recognise - to my knowledge - any category of grammatical relation between attributive adjectives and (non-copula) verbs. However, for verb-noun and adjective-noun collocations, the possibility exists of grouping the collocates automatically into grammatical relations, and this is what has been done in Tables 2-4.

The data displayed in these Tables (2 to 4) point to the existence of a well-defined class of noun collocates shared by *mental, occupational* and *cause*. One of the prevailing groups among the noun collocates of *mental* and *occupational* consists of words relating to *health* problems. This group includes nouns that denote a bad physical or psychological condition (e.g. *asthma, dermatitis, stress, deafness, disease, disorder, ill-health, distress, impairment, disability, incapacity, breakdown,* etc.) and nouns which refer to the field of medical care (e.g. *therapist, therapy, psychologist, psychology, physician, medicine, hospital, nursing,* etc.). The overlap with lexical selections found in the pattern caused *by* + NN is clear. In Table 4 we find several nouns referring to a bad physical or psychological condition (*infection, stress*), or to things and factors which can cause damage to physical or mental health (*smoking, parasite, virus, bacterium, trauma*).

therapist (10.81)	*physician (6.34)*	*medicine (4.95)*
therapy (8.45)	*hygienist (5.86)*	*functioning (4.93)*
asthma (8.24)	*psychology (5.84)*	*mobility (4.93)*
pension (7.43)	*stress (5.65)*	*grouping (4.85)*
exposure (7.41)	*ill-health (5.61)*	*hygiene (4.76)*
hazard (7.12)	*health (5.6)*	*disease (4.73)*
psychologist (7.06)	*competence (5.6)*	*profession (4.52)*
scheme (6.86)	*standard (5.49)*	*qualification (4.31)*
dermatitis (6.55)	*deafness (5.48)*	*School/Department (4.28)*
segregation (6.52)	*classification (4.95)*	*safety (4.21)*

Table 2. Word sketch of occupational (grammatical relation: modifier of).

illness (10.02)	*state (6.69)*	*disability (6.21)*
health (8.97)	*hospital (6.42)*	*capacity (6.13)*
disorder (8.69)	*note (6.4)*	*representation (6.12)*
problem (7.81)	*nurse (6.4)*	*breakdown (6.1)*
retardation (7.76)	*ill-health (6.37)*	*service (5.96)*
distress (7.69)	*incapacity (6.33)*	*faculty (5.84)*
well-being (7.1)	*arithmetic (6.26)*	*stimulation (5.83)*
impairment (6.95)	*professional (6.26)*	*trust (5.78)*
difficulty (6.86)	*nursing (6.24)*	*attitude (5.74)*
handicap (6.77)	*wellbeing (6.23)*	*agility (5.71)*

Table 3. Word sketch of mental (grammatical relation: modifier of).

negligence (7.86)	*deficiency (5.99)*	*fault (5.39)*
bacterium (7.44)	*defect (5.95)*	*trauma (5.38)*

virus (7.19)	*failure (5.87)*	*mite (5.36)*
mutation (6.91)	*misuse (5.75)*	*delay (5.36)*
lack (6.6)	*accident (5.68)*	*pollution (5.35)*
exposure (6.6)	*earthquake (5.57)*	*strain (5.28)*
infection (6.46)	*imbalance (5.49)*	*combination (5.27)*
smoking (6.4)	*error (5.44)*	*organism (5.25)*
fungus (6.09)	*stress (5.41)*	*asbestos (5.23)*
parasite (6.05)	*inflammation (5.41)*	*maladministration (5.19)*

Table 4. Word sketch of cause (grammatical relation: pp_by-i).

Basing on the information presented in Tables 2, 3 and 4, we can conclude that the expression *caused by occupational stress* in example (1a) can be analysed as a product of shared collocations of the noun *stress* with the verb *cause* and with the adjective *occupational*. The convergence of these two different collocations in a single expression is motivated by an overlap in the semantic selections made by the adjective *occupational* and by the verb *cause*. As stated, *stress* is a noun collocate shared by *occupational* and *cause* (see Tables 2 and 4), and it coheres with the characteristic semantic contexts of both these two words.

Likewise, we can conclude that the expression *caused by mental disorder* in example (1b) is a product of shared collocations of the noun *disorder* with the verb *cause* and with the adjective *mental*. We know that *disorder* is a collocate shared by *mental* (see Table 3) and *cause*. In this case, the shared collocate has not been listed in Table 4, but this does not mean that *disorder* does not form a statistically significant collocation with *cause*. As a matter of fact, it does (the score is 3.55), although it does not rank among the top 30 collocates.

In sum, examples (3a) and (3b) illustrate the characteristics of what can be called *direct inter-collocation*. This is a type of inter-collocational relation characterised by a unidirectional process of semantic selection. In this type of inter-collocational patterning, there is only one target of semantic selection for a combination of two different collocations (see Figure 1). Besides, the direction of semantic selection correlates with a grammatical aspect. The mechanism of direct inter-collocation requires that each of the two interlocking collocations be a word pair consisting of a predicate and the head constituent of the argument phrase (verb + noun, adjective + noun). In this sense, we can say that direct inter-collocation is facilitated by an isomorphism between the structure of lexical attraction and the syntactic structure of the argument phrase. This mechanism stands in sharp contrast to oblique inter-collocation, as I explain below.

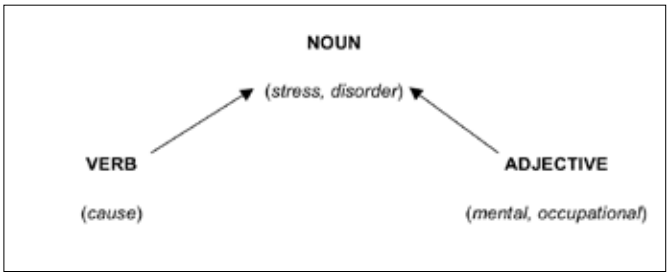

Figure 1. Direction of semantic selection in direct inter-collocation.

3.2 Analysing oblique inter-collocation

The adjectives *faulty* and *defective* are near-synonyms and, consequently, we expect their collocational behaviour to be very similar. An elegant way of showing these similarities is by resorting to *Sketch-Diff*, a SketchEngine tool which sorts the collocations of two lemmas of the same word class into two blocks: common patterns and exclusive patterns. The common patterns of *faulty* and *defective* include the words listed in Table 5. The nouns found in this list are statistically significant (logDice) items in head-modifier relations with these adjectives. The full list of common patterns is longer, but due to limitations of space only 30 items have been included (the selection is made automatically by the corpus query engine). Next to each word are given two different figures in brackets. They indicate the statistical scores of collocations with *faulty* and *defective*, respectively.

The collocations of these two adjectives are structured around a central notion of 'design'. There is a strong presence of nouns denoting products, i.e. objects which have been produced as a result of a manufacturing process (*tyre, brake, valve, heater, battery, hardware, lamp, machine*, *goods*, etc.). To this we must add the presence of nouns denoting the process itself of producing something following a particular design and skill (*manufacture, workmanship*). Finally, another group consists of nouns which denote small parts of organisms, or substances found in them (*protein, gene, chromosome*).

vision (0.5) (3.5)	*workmanship (7.3) (7.1)*	*battery (3.6) (2.1)*
tyre (2.5) (4.5)	*heater (3.5) (3.2)*	*device (1.8) (0.3)*
rail (1.2) (2.5)	*item (4.0) (3.5)*	*equipment (4.1) (2.4)*
protein (1.7) (2.9)	*component (3.1) (2.5)*	*valve (5.4) (3.6)*
brake (3.7) (4.5)	*gene (6.7) (5.9)*	*drive (2.2) (0.4)*
circuit (2.9) (3.6)	*unit (2.2) (1.2)*	*machine (2.1) (0.3)*
product (2.6) (3.1)	*hardware (2.9) (1.8)*	*memory (2.2) (0.4)*
pixel (3.6) (3.8)	*lamp (3.2) (1.9)*	*light (3.1) (0.7)*
chromosome (3.2) (3.4)	*copy (2.7) (1.4)*	*manufacture (4.8) (2.4)*
sensor (4.2) (4.1)	*good (6.0) (4.6)*	*appliance (5.8) (3.3)*

Table 5. Common patterns of faulty and defective (grammatical relation: modifier of).

Comparing Tables 4 and 5, we can observe that the collocation with *workmanship* and *goods* is much more characteristic of the adjectives *faulty* and *defective* than it is of the verb *cause*. The nouns *workmanship* and *goods* cohere with the habitual semantic contexts of *faulty*/*defective* but not with the negative semantic prosody of *cause*. It has to be conceded that *goods* is also a statistically significant co-occurrence of *cause*, however weak (logDice score: 0.6), but this pattern can be a result of what in the Lexical Constellation Model is known as *indirect collocation* (Cantos and Sánchez, 2001; Almela (forthcoming); Almela et al. (forthcoming)). The collocation of *goods* with *cause* is usually mediated by collocates that are shared by the two nodes, such as *damage, faulty,* or *defective*.

It follows that in examples (4a) and (4b), there are two parallel operations of semantic selection, each of them with a different source and a different origin. The semantic selection made by the verb *cause* affects the choice of the adjective (*faulty, defective*), and in turn, the semantic selection made by the adjective affects the choice of the noun (*workmanship, goods*).

A word of caution is in place here. With the above picture, I am not implying any hypothesis about the mechanism of psycholinguistic processing of these sentences. That is an entirely di-

fferent question. The course of a psycholinguistic process cannot be accessed on the sole basis of corpus data. The corpus provides us with a means for describing the way in which different language units relate to one another, but it provides us with no privileged access to the way in which they relate to the speaker. The division between psycholinguistics and corpus linguistics is sharp (Teubert, 2005).

Therefore, the above analysis of the directionality of semantic selection should not be interpreted as referring to the order or sequence in which collocations are processed in the speaker's mind/brain. It only refers to the structure of dependency relations among language units *as observed in the discourse*.

Going back to our analysis, the relations of lexical and semantic dependency among words observed in examples (4a) and (4b) diverge considerably from those observed in examples (3a) and (3b). In expressions such as *caused by faulty workmanship* and *caused by defective goods*, the process of semantic selection is oriented in a different direction for each of the two collocations. The verb selects the semantic class of the adjective, while the adjective in turn selects the semantic class of the noun. There is a different target of semantic selection for each of the two collocations involved in this pattern (see Figure 2). This can be described as a case of *oblique inter-collocation,* as opposed to *direct* inter-collocation. Oblique inter-collocation is associated with a lack of isomorphism between the structure of lexical attraction and the syntactic structure of the argument phrase. One of the two interlocking collocations consists of the combination of a predicate with a non-head constituent of the argument phrase.

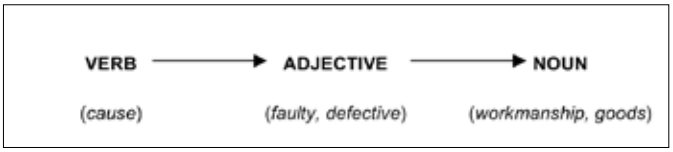

Figure 2. Direction of semantic selection in oblique inter-collocation.

Arguably, one of the contributing factors to oblique inter-collocation is an underlying strategy of *prosodic compensation*. The probability of finding adjectives of disapproval (*faulty, defective, incorrect*, etc.) after *caused by* depends to a large extent on the relationship between the noun and the semantic prosody of the verb. If the noun coheres with the negative prosody of *cause*, the likelihood of an adjective of disapproval occurring in this context increases substantially.

To test the empirical adequacy of this statement, I have measured the significance of associations between adjectives and specific realisations of the pattern *caused by + ADJ + NN* (the corpus and the association measure are the same ones as those used above). If the noun slot is filled in with the word *disorder*, the statistical significant adjectives are the following ones, in order of decreasing logDice score: *musculo-skeletal, digestive, concurrent, cerebral* and *genetic*. If the noun slot is filled in with the noun *stress*, the significant adjectives are *work-induced, oxidative, work-related* and *long-term*. Notice that of all the modifiers listed so far, only one (*oxidative*) expresses an intrinsically negative feature. In all other cases, the negative prosody of the verb is instantiated only in the combination with the noun.

Similar findings are obtained for other nouns which, like *stress* and *disorder*, contribute to the negative semantic prosody of *cause*. If the noun slot is filled with the word *illness*, the significant adjectives are *sudden*, *mental* and *long-term*. If the same slot is filled with *imbalance*, the significant adjectives are *biochemical, hormonal, chemical* and *nutritional*. Again, the conclusion is that the negative prosody of the verb is primarily conveyed by the noun, and not by the adjective.

In contrast, the choice of a noun which is not representative of the negative prosody of *cause* favours the choice of an adjective which is representative of that prosody (perhaps by way of compensation). Thus, if the noun slot after *caused by* is filled with the word *gene*, the significant adjectives are *recessive*, *faulty*, *defective*, *abnormal*, *inappropriate*, *altered*, *inherited* and *single*. Observe that half of the adjectives in this list (*faulty, defective, abnormal, inappropriate*) are also found in Table 1. Besides, they are representative of the negative prosody of *cause*, in the sense that they express an inherently negative evaluation of the referent of the noun they modify. If the noun slot is filled with *manufacture*, the list is shorter, but the results are similar: only the adjective *faulty* has statistical significance in this context. Of course, the results are also similar for the examples of inter-collocation analysed in (4a) and (4b). If the noun slot is filled with *workmanship* or *goods*, the only adjectives to reach the minimum score are *faulty* and *defective*.

This cannot be explained alone by the collocations of nouns and adjectives independently of the verb *cause*. The results differ substantially between the intra-collocational and the inter-collocational perspective. From a purely intra-collocational perspective, i.e. considered independently of the contexts shared with *caused by*, we can observe how the noun *goods* collocates with a wide range of pre-modifying adjectives, encompassing many semantic classes (e.g. *humanitarian, worldly, bulky, second-hand, perishable, counterfeit, dangerous, durable*, etc.). In abstraction from the collocations with *cause*, adjectives of negative evaluation such as *faulty* and *defective* form only a small subset of the adjectival collocates of *goods*. This subset becomes prominent in inter-collocation with *cause* and *goods*, but not in collocation with *goods* alone.

The conclusion arising from this evidence is that when the noun expresses the negative semantic prosody of the verb, it is unlikely to find this prosody reiterated in the adjective; and conversely, when the noun fails to express the negative prosody of *cause*, this function is very likely to be fulfilled by the adjective. This reinforces the notion of a mechanism of prosodic compensation. The mechanism tends to avoid both the lack of expression for the prosody of the verb and the repetition of the same prosody in two different constituents of the argument phrase. The verb *cause* induces the manifestation of its negative prosody in the agentive *by*-phrase, but it does not determine which actual constituent in particular must give expression to such prosody. In those cases in which the head constituent of the noun phrase fails to be representative of the habitual semantic associations of the verb, the flow of semantic selection is, so to say, diverted to a non-head constituent. Otherwise, the negative prosody of *cause* is reflected in the noun.

3.3 Direct and oblique collocation in the OCD

From a lexicographical perspective, patterns of direct inter-collocation are less problematic than oblique ones. Normally, the relevant information about the former can be traced through successive entries. For instance, the information necessary for analysing the expression *caused by occupational stress* is distributed over three different entries of the OCD. Firstly, the collocation of *cause* and *stress* can be found both in the entry for the verb and in the entry for the noun (see Figure 3). *Stress* is given as one of the subject collocates of *cause*, and *cause*, in turn, as one of the verbs that are typically combined with *stress*. Secondly, information about the collocation of *stress* and *occupational* is also found in two entries. *Occupational* is indicated as an adjectival collocate of *stress*, and *stress* is indicated as a noun collocate of *occupational* (see next page). Generally speaking, we can say that the structure of a conventional collocation dictionary, as is the OCD, is suitable for providing users with an access to patterns of direct inter-collocation.

In contrast, the phenomenon of oblique inter-collocation poses a challenge to the conventional way of organising collocational information in combinatory dictionaries. This phenomenon leads to the formation of co-occurrence patterns which are relevant for the knowledge of the words involved, but which do not fit well into any of the established syntactic types of collo-

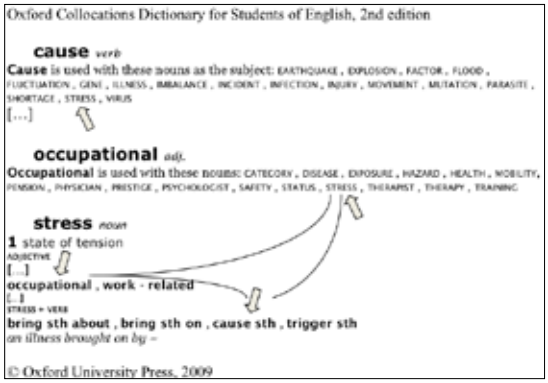

Figure 3. Tracing inter-collocations across OCD entries.

cation. As we have seen, the collocation of *cause* with *faulty* and *defective* is as much representative of the usage of *cause* as is the collocation of this verb with nouns such as *disorder* or *stress*, among others. Yet, the information which collocation dictionaries provide about this type of verb-adjective patterning is only fragmentary and incomplete, compared to the systematicity and the abundance of information they provide about verb-noun collocations.

Observe, for instance, that the OCD offers only a part of the information needed for analysing expressions such as *caused by defective goods* and *caused by faulty workmanship*. In these cases, the dictionary offers information concerning the collocation of the adjective and the noun. The user is informed that *goods* is a collocate of *defective* and that, in turn, *defective* is a collocate of *goods*. Likewise, we can find the noun *workmanship* in the entry for *faulty*, and *faulty* in the entry for *workmanship*. However, this information accounts for only a part of the inter-collocational pattern.

My objection is not that *goods* and *workmanship* fail to appear in the entry for *cause* (and vice versa). This decision is well justified, considering that none of these two nouns is a typical collocate of *cause* (see section 3.2). Rather, my objection is that there is no information concerning the collocational connection between *cause* and *faulty*/*defective*. None of these adjectives is recorded in the lexical entry for *cause*, and conversely, *cause* is not recorded in the entries for any of these adjectives.

This decision is less justified, not only because *faulty* and *defective* are statistically significant co-occurrences of *cause*, but also, and more importantly, because they are good examples of the semantic selection made by the verb.

4. CONCLUDING REMARKS

In this article I have argued that the negative semantic prosody of the verb *cause* is expressed in collocations of varying syntactic forms, including structures which are not contemplated in the established typologies, as is the case of collocations of transitive verbs and attributive adjectives (e.g. *cause by + faulty/defective + NN*). I have also argued that this kind of syntactically atypical collocations of *cause* follows a specific strategy of interaction with other collocational patterns that co-exist in the same phrase. The choice of an adjective is affected by whether, or not, the noun coheres with the negative semantic prosody of the verb. In those cases in which the noun does not express the semantic selection made by the verb, the adjective is coerced into doing

so. This mechanism of *prosodic compensation* is characteristic of oblique inter-collocations, i.e. of combinations of collocational pairs which share one component but have different targets of semantic selection.

It is the task of future research to determine the extent to which the examples analysed here are representative of a broader phenomenon in the language. The conclusions I have drawn from this study apply only to inter-collocations of *cause* with nouns and adjectives. From a single case study we cannot arrive at definite conclusions about the principles governing semantic selection and lexical attraction in patterns of inter-collocation. So far, the findings are preliminary. However, they open up a question for discussion and highlight a path for further inquiry. The characteristics observed in inter-collocational patterns of *cause* cannot be adequately accounted for by the existing models of collocation, and the implications for collocation dictionaries should not be understated.

At present, the OCD can only partially respond to the demands of a user who seeks information about the contexts of use of the verb *cause*. Collocations with adjectives such as *faulty* or *defective*, among others, are highly representative of the usage and the meaning of this verb, but they will not be found in the dictionary. This gives us good reasons for further inquire about possible alternations of direct and oblique inter-collocations with other verbs. Future research will decide whether the interaction observed among collocations of *cause* is an isolated phenomenon, or whether it reveals a more general problem, namely the obsolescence of the received syntactic typologies of collocation, with the consequent need to introduce modifications in the micro-structure of collocation dictionaries.

REFERENCES

Almela, M. (2006). *From Words to Lexical Units. A Corpus-Driven Account of Collocation and Idiomatic Patterning in English and English-Spanish*. Frankfurt: Peter Lang.

Almela, M. (forthcoming). "Improving Corpus-Driven Methods of Semantic Analysis: A Case Study of the Collocational Profile of *Incidence*". *English Studies*.

Almela, M., P. Cantos & A. Sánchez (forthcoming). "From Collocation to Meaning: Revising Corpus-Based Techniques of Lexical Semantic Analysis", in I. Balteiro (ed.) *New Approaches to Specialized English Lexicology and Lexicography*. Newcastle: Cambridge Scholars, 47-62.

Alonso Ramos, M. (2001). "Construction d'une base de donées des collocations bilingue français-espagnol". *Langages* 143: 5-27.

Alonso Ramos, M. (2007). "Actantes y colocationes", *Nueva Revista de Filología Hispánica* 2: 435-458.

Apresjan, J. (2009). "The Theory of Lexical Functions: An Update", in D. Beck, K. Gerdes, J. Milićević & A. Polguère (eds.) *Proceedings of the Fourth International Conference on Meaning-Text Theory. Proceedings*. Montréal: OLST, 1-14.

Barrios, M.A. (2009). "Domain, Domain Features of Lexical Functions, and Generations of Values by Analogy according to the MTT Approach", in D. Beck, K. Gerdes, J. Milićević & A. Polguère (eds.) *Proceedings of the Fourth International Conference on Meaning-Text Theory. Proceedings*. Montréal: OLST, 45-53.

Barrios, M.A. (2010). "El dominio de las funciones léxicas en el marco de la Teoría Sentido-Texto", *ELiEs* 30*,* in *http://www.*elies.rediris.es/elies30/ [Date of access: 30.11.2010]

The BBI Dictionary of English Word Combinations (1997). Compiled by M. Benson, E. Benson & R. Ilson. Amsterdam: John Benjamins.

Bosque, I. (2001). “Sobre el concepto de ‘colocación’ y sus límites”, *Lingüística Española Actual* 23-1: 9-40.

Bosque, I. (2004). “La direccionalidad en los diccionarios combinatorios y el problema de la selección léxica”, in T. Cabré (ed.) *Lingüística teórica: anàlisi i perspectives*. Bellaterra: Universitat Autonoma de Barcelona, 13-58.

Cantos, P. & A. Sánchez (2001). “Lexical Constellations: What Collocates Fail to Tell”, *International Journal of Corpus Linguistics* 6-2: 199-228.

Corpas Pastor, G. (1996). *Manual de fraseología española*. Madrid: Gredos.

Corpas Pastor, G. (1998). “Expresións fraseolóxicas e colocacións: clasificación”, in *Actas do I Coloquio Galego de Fraseoloxía* 31-61.

Corpas Pastor, G. (2001). “Apuntes para el estudio de la colocación”, *Lingüística Española Actual* 23-1: 41-56.

Cruse, D. (2004). *Meaning in Language. An Introduction to Semantics and Pragmatics*. Oxford: Oxford University Press.

Hausmann, F. (1998). “O diccionario de colocacións. Criterios de organización”. In *Actas do I Coloquio Galego de Fraseoloxía* 63-81.

Hausmann, F. (1979). “Un dictionnaire des collocations est-il posible?”, *Travaux de linguistique et de literature* 17-1: 187-195.

Luzón, M.J. & M. C. Campoy (2000). “The collocation of linking verbs of transition and adjectives”, *Cahiers de Lexicologie* 76: 41-55.

Martin, W. (2008). “A unified approach to semantic frames and collocational patterns”, in S. Granger & F. Meunier (eds.) *Phraseology. An interdisciplinary perspective*. Amsterdam: John Benjamins, 51-65.

Mel’čuk, I., A. Clas & A. Polguère (1995). *Introduction a la lexicologie explicative et combinatoire*. Louvain-la-Neuve: Duculot.

Oxford Collocations Dictionary for Students of English (2009). Compiled by C. McIntosh. Oxford: Oxford University Press.

REDES. Diccionario combinatorio del español contemporáneo (2004). Compiled by I. Bosque. Madrid: SM.

Rychlý, P. (2008). “A Lexicographer-Friendly Association Score”, in P. Sojka & A. Horák (eds.) *Proceedings of Recent Advances in Slavonic Natural Language Processing*, RASLAN 2008, Brno: Masaryk University, 6-9.

Stubbs (1995). “Collocations and semantic profiles. On the cause of trouble with quantitative studies”, *Functions of Language* 2-1: 23-55.

Stubbs (2002). *Words and Phrases. Corpus Studies of Lexical Semantics*. Oxford: Blackwell.

Teubert, W. (2005). “My version of corpus linguistics”, I*nternational Journal of Corpus Linguistics* 10-1: 1-13.