JOURNAL OF ENGLISH STUDIES – VOLUME 17 (2019), 3-41. http://doi.org/10.18172/jes.3905

COLLOCATION AND SELECTIONAL PREFERENCES: A FRAME-BASED APPROACH

MOISÉS ALMELA-SÁNCHEZ
Universidad de Murcia
moisesal@um.es

ABSTRACT. *Most of the research conducted into collocation and semantic frames has dealt with these phenomena separately. The study of collocation has not figured prominently in the research agenda of frame semantics, and frame semantics has only sporadically been used as an analytical framework for collocation. This article is a contribution to narrowing the gap between the two fields. It does so by addressing key issues in the design of a frame-based approach to collocation, with a special focus on the relation between collocational patterns and semantic valency, and by providing arguments for the efficacy of the frame-semantic theoretical apparatus in explaining verb-adjective links that are not accounted for by the existing models of collocation. The methodology combines lexicographic resources as well as quantitative and qualitative analysis of examples and data from an English web corpus (ukWaC).*

Keywords: Lexical semantics, corpus linguistics, semantic frames, collocation, valency, cognitive grammar.

LA COLOCACIÓN Y LAS PREFERENCIAS DE SELECCIÓN: UN ENFOQUE BASADO EN CORPUS

RESUMEN. *Las investigaciones sobre colocaciones y marcos semánticos han discurrido, en su mayor parte, por caminos separados. Ni las colocaciones ocupan un lugar prioritario en los estudios sobre marcos ni es habitual aplicar la semántica de marcos como enfoque de estudio de las colocaciones. Este artículo aspira a estrechar la relación entre ambos campos. Para ello, abordaré una serie de cuestiones clave para el desarrollo de un enfoque de investigación sobre colocaciones basado en marcos. Me centraré, sobre todo, en la relación con la valencia semántica. Además, el artículo aporta argumentos que justifican la eficacia del enfoque basado en marcos para explicar un tipo de vínculo semántico entre verbos y adjetivos que escapa a los actuales modelos descriptivos de colocaciones. La metodología aplicada combina la utilización de recursos lexicográficos con el análisis cuantitativo y cualitativo de ejemplos y datos de uso extraídos de un corpus de inglés basado en la web (ukWaC).*

Palabras clave: semántica léxica, lingüística de corpus, marcos semánticos, colocación, valencia, gramática cognitiva.

Received 5 December 2018
Revised version accepted 7 May 2019

1. INTRODUCTION

Collocation studies and frame semantics are mutually relevant areas of research. The former deals with lexical syntagmatic patterns, and the latter with representations of stereotyped states of affairs. The common ground between the two fields is based on two aspects: both claim a prominent role in the research agenda of empirical semantics, and both are interested in analysing expectation patterns evoked by individual lexical items.

In the existing literature, the connection of collocation and frames has been addressed with four main goals:

- deriving collocational information from frame-based lexical resources, in particular from FrameNet (Ruppenhofer, Baker and Fillmore 2002; Alonso Ramos, Rambow and Wanner 2008);
- using collocational data from existing lexical resources or from corpora for descriptions of semantic frames (Fontenelle 2009; Johnson and Lenci 2011; Akita 2012; Dalpanagioti 2013; Buendía Castro, Montero Martínez and Faber 2014; Buendía Castro and Faber 2017; Montero Martínez and Buendía Castro 2017);
- combining collocational information and semantic frame descriptions in a lexical resource, as in the DiCoEnviro specialised dictionary (L'Homme, Laneville and Azoulay 2014);

– applying a frame-semantic theoretical apparatus to the analysis of collocational phenomena (Martin 2003, 2008).

The goal of this article fits into the fourth group. Like Martin (2003, 2008), I will take advantage of frame-based theoretical notions for improving the analysis of collocational patterns. However, there is a fundamental difference concerning the specific content of the proposal. While Martin's approach sets the analysis of collocational patterns in relation to the phraseological spectrum, the approach I adopt here emphasises the relation between collocation and valency patterning. The main goal is to argue for the efficiency of frame semantics – and of its Fillmorean branch in particular – as a theoretical framework for explaining and analysing the patterns of semantic selection observed in the collocational range of predicative lexemes. I will do that with special emphasis on problems arising from the analysis of semantic selection in non-head components of argument slots.

2. PERSPECTIVES ON COLLOCATION

There is no widely shared notion of collocation. Most experts agree that collocational phenomena form co-occurrence patterns among words, but there is no consensus on which filters should be imposed on collocation candidates. Three main types of requirements can be distinguished:

– Quantitative filters: the co-occurrence must be sufficiently frequent and statistically relevant according to a specific measure of lexical association applied to corpus data (Jones and Sinclair 1974; Mason 2000; Stubbs 2002; Gries 2013).
– Qualitative filters (I): syntactico-semantic dependency. The standard typology includes the following patterns: V+N, Adj+N, Adv+V, Adv+Adj, N+Prp+N (Corpas Pastor 1996; Hausmann 1998; Martin 2008). This typology is informed by the idea that collocational relations operate within the structural framework of predicate-argument relations (Bosque 2001a, 2004a, 2004b; Tutin 2008).
– Qualitative filters (II): phraseological binding. In the realm of collocations, this property is attributed to constraints operating in the opposite direction to canonical valency patterning and selectional restrictions, i.e. from argument to predicate (Írsula Peña 1994; Hausmann 2007; Alonso Ramos 2007). The dominant or autonomous element in this relationship (the argument) is called the *base*, and the dependent element (the predicate) is called the *collocator* or *collocate*[1] (Mel'čuk 1998; Martin 2008).

[1] This concept should not be confused with the notion of *collocate* in Firthian linguistics, where it denotes any statistically relevant co-occurrence of a search item (the *node*).

Different approaches have opted for different combinations of filters. The two dominant approaches so far, often identified under the labels of *British contextualism* and *Continental* tradition (Williams 2003; Siepmann 2005; Tutin 2008), have privileged the quantitative and the qualitative filters, respectively. The main exponent of the former is the Sinclairian branch of corpus linguistics (Sinclair 1991); the second one has been mainly inspired by the theories of Hausmann (1979, 1997) and Mel'čuk (1998, 2003). Although these two authors developed their research separately – one as a leading scholar in European lexicography, the other as one of the fathers of Meaning-Text Theory (MTT) – the similarities between their respective insights into collocation are profound and amply recognized (Heid 1994; Mel'čuk, Clas and Polguère 1995).

Some studies combine quantitative and qualitative categories of description. This is done from two different perspectives and with two different goals:

- enriching a phraseological approach to collocation with frequency information (Koike 2001; Vincze and Alonso Ramos 2013);
- improving the operationalisation of the statistical notion of collocation through the incorporation of syntactic specifications (Kilgarriff and Tugwell 2001; Nerima et al. 2010; Seretan 2011; Uhrig and Proisl 2012; Bartsch and Evert 2014).

In this study, I will use syntactic filters with a view to refining the output of collocation statistics. The motivation for combining these two types of filters is the possibility to distinguish two different aspects or dimensions of collocational phenomena. First, from a purely empirical standpoint, a collocation manifests itself in the combinatory bias of particular words in language use. The directly observable property of collocational patterns is the tendency of individual words to privilege particular lexical contexts and to avoid others. At present, lexical association measures represent the most effective and reliable means of capturing this aspect of collocation (i.e. what Evert (2009) calls the "primary data"). Second, from a descriptive standpoint, one of the most relevant characteristics attributed to statistical collocational patterns is their tendency to be composed of elements in direct syntactic relation. As Bartsch and Evert (2014: 60) report, "collocations have been shown by some studies to tend to form grammatical relations".

The third type of filter (phraseological binding of the collocator/predicate to the base/argument) is not applied in this study as a necessary condition for collocation status. In the specialised literature, this aspect of collocation has been an object of theoretical controversy. According to some authors, the phraseological character attributed to collocation in the theories of Hausmann and Mel'čuk results, at least in part, from the adoption of a perspective of analysis which is oriented from arguments to predicates (Bosque 2001a, 2004b, 2017; Apresjan and Glovinskaja 2007; Apresjan 2009; for a response to this objection, see Alonso Ramos 2017). When collocational

patterns are analysed from the opposite point of view, i.e. from the standpoint of the conditions that predicates impose on their argument fillers, the semantic motivation and the systematicity of collocations become more noticeable.

This criticism of the Hausmann/Mel'čuk conception of collocation can be illustrated with the following example. Consider the collocations *shoot a picture* and *sit an exam*. If they are analysed from the point of view of the noun, what stands out are the differences in the selection of specific collocators, e.g. *make/take/shoot/*sit a picture* vs. *make/take/sit/*shoot an exam*. The selection of a verb for expressing the meaning 'do, perform' is not independent of the context provided by the noun. However, if the same collocational pattern is approached from the perspective of predicate-driven constraints, what is brought out is the systematic selection of nouns from specific semantic sets. Thus, *shoot* does not collocate individually with *picture* but with a set of nouns denoting 'recording of (an) image(s)', e.g. *photo, photograph, video, film, movie, documentary*, etc. Likewise, the bond between *sit* an *exam* is not an item-specific (idiosyncratic) link but rather forms part of a broader combinatory pattern including other nouns from the same semantic type, e.g. *examination, test, A-level*, etc. Thus, from the perspective of the predicate, collocational patterns appear as specific lexical realisations of more schematic patterns of semantic selection. This perspective shifts the study of collocation away from the realm of phraseology and brings it in closer relation to semantic valency.

Finally, another question that needs to be addressed concerns the nature of the restrictions or preferences imposed on argument classes. Bosque (2001a, 2004a, 2017) and Sánchez Rufat (2010) draw a clear-cut distinction between linguistically motivated and non-linguistically motivated constraints. However, as Írsula Peña (1994) observed in relation to Coseriu's (1977) notion of *lexical solidarities*, the distinction between the two types of combinatory constraints ultimately reflects the *range* of the argument class, rather than the nature of the selection. Preferences for highly specific argument classes tend to be characterised as *intra-linguistic*, while broadly defined categories tend to be relegated to the *extra-linguistic*. My stance on this issue is aligned with the generalised view in frame semantics. I will not assume the plausibility of a strict qualitative divide between purely linguistic and non-linguistic factors of predictability in lexical combinatorics.

3. COLLOCATIONAL DATA AND FILLER INFORMATION

Among the different trends in frame semantics, the one that lays greater emphasis on valency[2] description is the one associated with the work of Fillmore

[2] In the literature, there is an alternation between the terms *valency* and *valence*. Roughly speaking, the former is the European variant, and the latter is the American variant.

and the FrameNet project – for a comparison with other approaches to frames, see Busse (2012). The hallmark of the FrameNet approach is the integration of valency patterns within an onomasiological approach to word meaning (Boas 2001; Fillmore 2007; Baker 2009). In this framework, conceptual representations of scenes or states of affairs are established as *definienda* (Boas 2001), and the combinatory potentialities of words which evoke such frames are analysed as a linguistic realisation of semantically relevant background information.

However, not all layers of valency patterning have received the same amount of attention in the FrameNet database and in the literature about this lexicographic project. The linking of semantic roles and syntactic forms has been a priority, while the description of semantic types of argument fillers has occupied a secondary position. Fillmore (2007) identified this gap and entrusted the task to later research:

> One property of a valency description which FrameNet has not managed to provide directly is an account of the typical semantic types of the phrases that serve as frame elements. It is hoped that later research based on further corpus evidence can spot the semantic types found for particular FEs of particular LUs and incorporate such results in the valency descriptions – beyond such limited high-level indications as animate, concrete, and abstract. (Fillmore 2007: 154)

This quotation combines two ideas with implications for future developments in frame semantics: the first one is the proposal to search for finer-grained semantic categorisations of frame element fillers; the second one concerns the methodology for accomplishing this goal. Regarding the first issue, it should be noted that what Fillmore describes in the above quote as "high-level indications" is, mutatis mutandis, equivalent to the notion of *selectional restrictions* in the generative jargon. Presumably, some frame elements (FEs) allow for more specific categorisations of their typical lexical fillers. Although experts are not unaware of the gap, a comprehensive lexicographic coverage of the phenomenon has not been accomplished yet. The function of semantic types in FrameNet is explained in Ruppenhofer et al. (2016), but descriptions of these features in recent releases of the database are, for the most part, still limited to highly schematic glosses (e.g. 'physical entity', 'sentient', 'state of affairs', etc.). In most entries, FrameNet has not incorporated yet precise semantic characterisations of FE fillers such as those provided for argument slots in other landmark lexicographic projects, particularly in the REDES dictionary of lexical restrictions in Spanish (Bosque 2004a), and in valency dictionaries of English and German (Herbst, Heath, Roe and Götz 2004; Schumacher, Kubczak, Schmidt, and de Ruiter 2004).

The second idea that needs emphasising in this respect concerns the methodology for identifying semantic types. As Fillmore predicts in the above

quote, the input of corpus linguistics can be decisive for accomplishing this task. The concept of collocation – along the lines defined in the previous section – fits readily into this programme. Some corpus studies indicate that the more representative fillers of FE slots tend to be grouped around specific semantic classes or types. This aspect of the combinatory potential of frame-bearers has received the name of *selectional preferences* (Johnson and Lenci 2011) or *collocate types* (Dalpanagioti 2013). Moreover, to the extent that these combinatory properties are shared by evokers of the same scene or situation, they can also be attributed to realisations of general properties of the frame:

> ...we believe that both the selectional preferences of LUs belonging to the same frame and the generalizations that can be drawn from these on the 'selectional preferences' of the entire frame are essential for defining the semantics of the frame itself and should be integrated, ideally, into the LU and FE definitions inside the FrameNet database. (Johnson and Lenci 2011: 19)

Although Johnson and Lenci's study (2011) is centred on methodological issues in semantic frame description, their conclusions point to possible theoretical implications. If, inductively, the selectional properties of frame-evoking items can be used as an empirical basis for identifying semantic properties at the level of frame structure, then, in a deductive move, it should also be possible to establish the organising principles of frame structures as a theoretical framework for explaining and analysing the distribution of collocate types in argument fillers. In the following section I will deal with this theoretical dimension of the relationship between collocational patterns and frame structures. The theoretical bases of the proposal are laid out in the next section. Section 5 shows how this can be applied in the analysis of collocational phenomena that resist a neat classification in other models of collocation.

4. CONCEPTUAL DEPENDENCY AND FRAME STRUCTURE

The cornerstone of the proposal presented here is an elaboration of the notion of *autonomy-dependency asymmetry* (abbreviated to *A/D asymmetry*), which Langacker considers "an essential feature of language design" (Langacker 1991: 286). In the next subsections I will offer a brief description of two different aspects or dimensions of A/D asymmetry as well as of their impact on frame structure.

4.1. FIRST AXIS: A/D ASYMMETRY AND FRAME EVOCATION

In cognitive grammar, canonical valency relations are portrayed as asymmetrical correspondences between substructures of a conceptually dependent element

and a conceptually autonomous element that elaborates (i.e. fills out details of) the former (Langacker 1987, 2002). The difference between the dependent and the autonomous element is determined by the impact of external relations on the internal structure of the concept.[3] The dependent structure "requires for its conceptualization some intrinsic reference – however schematic – to the entities that participate in the relation" (Langacker 1987: 300). This criterion allows Langacker to formulate some generalisations about prototypical classes of each type. As a general rule, relational notions are conceptually dependent, while conceptions of physical objects are prototypically autonomous elements.

These generalisations are qualified by the observation that valency relations can occur in non-canonical configurations. Besides, in consonance with the epistemology of cognitive linguistics, the autonomy-dependency distinction is conceived as a gradient rather than as a binary divide (Langacker 1987, 2002). Two components of a valency relation may exhibit some degree of conceptual dependency on the other, though to a different extent. For instance, as Sullivan (2013) observes, the semantic structure of man is not devoid of substructures that can be elaborated by the adjective *tall*, but the impact of the combination on the semantic structure of *tall* is greater, because "the concept of HEIGHT is meaningless without a referent that can be tall or short" (Sullivan 2013: 30-31). Nevertheless, with these caveats in mind, Langacker insists that in prototypical configurations of valency, the greater impact of the argument on the conceptualisation of the predicate justifies a qualitative distinction: "The dependent structure can be equated with the predicate, in predicate-argument terms, and the autonomous structures with its arguments" (Langacker 2002: 170).

Interestingly, Sullivan (2013) argues that A/D relations can be represented using semantic frame models. As she observes, the function of the autonomous element in the substructure of the dependent element is analogous to the function of a filler assigned to a FE. Both serve to fill out details of a substructure in a relation evoked by a conceptually dependent expression (a frame evoker):

> We can say, then, that the structure of a dependent element is specifically a frame structure, and that its elaboration site is a frame role. The structure of the autonomous element is a filler for this frame role, and elaboration itself consists of filling a frame role. Autonomy, dependence, and the relation between them can all be described with these frame semantic terms. (Sullivan 2013: 32)

[3] The notion of *semiotaxis*, formulated by Hausmann (1997, 1998, 1999), shows profound analogies with the Langackerian concept of A/D asymmetry, given that both classify co-occurring items according to the influence which their combinatory potential exerts on their semantic structure. An exhaustive discussion of the common ground and the differences between the two notions is beyond the scope of the present article.

This suggests that conceptual dependency is a characteristic condition of frame-evoking potential, because the capacity of a word for evoking a frame is in direct relation to the impact that the frame exerts on the meaning of the word. In fact, in FrameNet, words with intrinsically relational meanings – particularly verbs – are attributed a prominent role as evokers of frames (Fillmore, Johnson and Petruck 2003). Again, this is not to deny that words of other classes can also exhibit a frame-evoking potential, but, crucially, this potential is in direct relation to the presence of predicative properties. Thus, among nouns, there are substantial differences in frame-evoking capacity across types: "event nouns and relational nouns are most clearly frame-evoking" (Ruppenhofer et al. 2016: 43).

The above considerations suggest the following: (i) the relationship between a frame-bearing unit and the FEs evoked by it reflect features of A/D asymmetry; (ii) conceptual dependency can be considered as a contributing factor to valency-bearing and to frame-evoking potential.

4.2. SECOND AXIS: A/D ASYMMETRY IN FRAME-TO-FRAME RELATIONS

In the second volume of the Foundations of Cognitive Grammar, Langacker distinguished a second axis of A/D asymmetry, which is manifested in unidirectional relations of conceptual dependency between event components. This aspect of A/D asymmetry leads to A/D layering, where a sequence of increasingly complex events is built on less complex ones. This is a typical feature of events involving the exertion of volitional control or instances of perception. Langacker (1991) illustrates this with the following examples:

(1) a. The ice cracked.
b. A rock cracked the ice.
c. A waiter cracked the ice with a rock.
d. The manager made a waiter crack the ice with a rock.
e. The owner had the manager make a waiter crack the ice with a rock.

Example (1a) provides the nucleus for the layering which unfolds through examples (1b) to (1e). The change of state undergone by the PATIENT (The ice) in (1a) is conceptualised independently of other events, but each of the events characterised in the other examples presupposes this change of state. The assembly of events increases its complexity progressively as new layers of causation are added to the chain. This is represented in the following notation, taken from Langacker (1991: 292), where T stands for thematic relationship (i.e. a conceptually autonomous relationship involving a single participant) and E stands for event:

$$(T) > (E_1(T)) > (E_2(E_1(T))) > (E_3(E_2(E_1(T)))) > (E_4(E_3(E_2(E_1(T)))))$$

The progressive increase in the complexity of the event through successive layers of causation involves also the addition of successive participants. With each new layer of causation, an additional participant is portrayed as inducing the less complex event. Such participants can be described at various levels of schematicity. At their most schematic, they correspond to archetypal roles (in Langacker's terminology); at a finer-grained level of description, they correspond to frame-specific semantic roles (i.e. FEs).

Again, as in the description of canonical valency, it is possible to find a frame-semantic counterpart for the properties of A/D asymmetry. The concept of frame-to-frame relations (or, simply, frame relations) denotes a feature of the design of FrameNet which shows close correspondences with the key properties of A/D layering. A frame relation is defined as "a directed (asymmetric) relation between two frames, where one frame (the less dependent, or more abstract) is called the Super_frame and another (the more dependent, or less abstract) is called the Sub_frame" (Ruppenhofer et al. 2016: 79). The correspondences with properties of A/D layering are granted by three characteristics. The first one refers to the composition of the relations: like A/D layering, frame relations involve two or more sets of structured relations between semantic roles. The second characteristic is the asymmetric nature of the relation: the dependency of the subordinate frame on the superordinate frame[4] has no equivalent in the reverse direction (i.e. the superordinate frame is not dependent on the subordinate frame). Finally, the third characteristic refers to the relation between schematicity and conceptual autonomy: both in A/D layering and in frame relations, conceptual structures that are more dependent are also more specific (i.e. they include more components) than less dependent conceptual structures.

5. VALENCY LAYERING AND STRATIFIED COLLOCATION: TWO CASE STUDIES

In this section, I will apply the properties of A/D asymmetry to the analysis of patterns of semantic selection involving frame relations. The first axis of A/D asymmetry establishes that frame-evoking predicates provide a schematic characterisation of FE fillers, and the second axis predicts that frames which represent conceptually dependent events contain structures of less dependent

[4] In FrameNet, the term *subframe* is also used to refer to a specific type of frame-to-frame relation holding between a sequence or phase in a complex event and the representation of the whole event (for example, between Arrest and Criminal_process) (Ruppenhofer et al. 2016). To avoid terminological ambiguity, I will distinguish this specific type of frame relation from the more general notion of conceptual dependency of less abstract events towards more abstract events, situations, or states of affairs. The terminological pair of *subordinate/superordinate* can serve to describe this generic notion of dependency.

(less complex) event representations. A corollary of this is that the selectional preferences of predicates evoking a subordinate frame will contain predicates evoking a superordinate frame. As I will argue, this allows us to predict patterns of distribution of selectional preferences which cannot be accounted for by the mainstream models of collocation analysis. I will illustrate this with reference to two cases studies.

5.1. METHODOLOGY

The methodology applied in this study combines the top-down procedure and the qualitative analysis characteristic of the FrameNet methodology with a bottom-up approach featuring quantitative analysis of corpus data. In this respect, the methodology is similar to that of Johnson and Lenci (2011), particularly in what concerns the use of collocational data for analysing selectional preferences of target lexical units and of the frame structures they evoke. There is, however, a slight difference in the objectives that motivates also a partial difference in the strategies employed. In the present study, the analysis of collocate types is oriented to the description of relations of conceptual dependency between frames of different levels of complexity. For this purpose, a specific two-step methodology has been devised.

The first step of the methodology is centred on qualitative, conceptual analysis of frame structures and frame relations. The goal of the analysis at this stage is to identify possible relations of A/D layering between semantic frames which overlap in their internal structure but have different levels of internal complexity. This step of the methodology is carried out with the aid of lexicographic information from FrameNet (https://framenet.icsi.berkeley.edu) complemented with qualitative analysis of concordances from a corpus. The corpus used for obtaining example concordances is ukWaC (Ferraresi, Zanchetta, Baroni and Bernardini 2008). This is a web corpus of British English containing 1,313,058,436 tokens. The corpus can be accessed through the Sketch Engine query system. There are bigger corpora of English, particularly those of the enTenTen family, but on balance, the ukWaC offers the advantage of being more homogeneous.

The second step involves the analysis of selectional preferences through collocational data. This step balances a quantitative and a qualitative dimension. The quantitative aspect consists in the application of collocation statistics for extracting typical slot fillers, and the qualitative aspect consists in the description of semantic sets of collocates. The corpus used for extracting the data is the same as in the first step (ukWaC), and the score used for measuring the strength of lexical association is logDice (Rychlý 2008). Only collocates reaching the threshold

of statistical significance are taken into account. In the case of logDice scores, the threshold is 0, so that, in principle, all collocates with positive scores can be included. Based on the criteria explained in Section 2, I adopted a syntactic-relational approach to the definition of the collocational search space. Collocates are searched in specific syntactic slots connected to the node (the search item). Where space constraints advised limitation of the maximum number of items, the criterion used for the cut-off point was the association score.

The syntactic pattern selected was: V+Adj+N. In the case of forms of participial origin, such as *hidden*, *protected*, *astonishing*, *speeding*, *voting*, *unpublished*, etc., the criterion for assigning them adjective status was based on lexicographic information. Only those that are registered as adjectives in a separate entry in the Oxford Dictionary of English (ODE) were included in the list. A similar criterion was applied to the treatment of compounds. Thus, *arrestable* and *ethnic* were excluded from the list of collocates of *commit* and *perpetrate*, respectively, because in all the co-occurrences with these verbs, these adjectives occur as part of expressions (*arrestable offence*, *ethnic cleansing*) for which the ODE reserves a separate entry as nominal compounds. The same negative filter was applied to *cardinal* (*commit* + *cardinal sin*).

The procedure for extracting the collocates and obtaining the association scores was divided into three steps. First, using the Word Sketch tool, I accessed the concordances for all the "object_of" relations of each search item (a verb). Then, using the Filter tools and the Corpus Query Language, I selected all those concordances in which the verb co-occurs with an object noun (this also includes subject nouns in the passive) modified by an adjective. Finally, using the Collocation tool, I obtained the inventory of collocates and their logDice scores. The results automatically obtained from the corpus tools were manually revised in order to filter out non-relevant occurrences (for instance, other verbs or nouns co-occurring with the target V+Adj+N patterns). Where a word in the collocate list was grammatically ambiguous (e.g. *material*), the logDice score was recalculated using the frequency data of the adjective (this recalculation was performed by means of an Excel file).

One obstacle to the methodology applied here is the polysemy of target words. In some cases, the semantic range of a collocate inventory reflects the variability of the node word's meaning rather than the range of semantic types ascribed to a single frame. Unfortunately, this difficulty cannot be reliably overcome in the phase of automatic collocation extraction. At present, there is no means of automatically assigning concordances to frames. A subcorpus which contains only concordances of the relevant semantic frame cannot be built on a large scale, and consequently, the collocates must be extracted for occurrences of the target word regardless of frame assignment. This difficulty can be dealt with in a later phase

of the methodology, particularly through the interpretation and the qualitative analysis of collocates and example concordances.

Finally, another methodological decision concerns the lexical resources used for ascribing words to frames. The FrameNet Project is still in the making and it makes no claim of exhaustivity for its database. It should not be assumed that the words listed as target lexical units under the entry for a given frame represent the entire stock of words in English for which that frame provides a relevant conceptual background (thus, *secret*, *confidential*, and *unclassified* are documented under the SECRECY_STATUS frame, but *classified* is not). Where necessary, the use of FrameNet data was complemented with information from a thesaurus –the Oxford Thesaurus of English (OTE)– and from general-purpose dictionaries: the aforementioned ODE, the Longman Dictionary of Contemporary English (LDOCE), and the Macmillan English Dictionary (MED). The use of these resources in the present study was guided by the same criteria that generally inform the grouping of words into frames in FrameNet (Ruppenhofer et al. 2016: 11-17), with special attention to the following: similarity of semantic type, profiling of the same frame elements, and near-paraphrasability. The first criterion establishes that "the basic denotation of the targets in a frame should be similar" (Ruppenhofer et al. 2016: 14). The second criterion (same profiling of frame elements), implies that "the same participant's point of view should be emphasized" with all the lexical units from a frame (Ruppenhofer et al. 2016: 13). Finally, the criterion of near-paraphrasability is met when "one can more or less felicitously substitute one lexical unit for another and still evoke the same frame and express the same kinds of semantic roles as syntactic dependents of the new lexical unit" (Ruppenhofer et al. 2016: 15). Generally speaking, quasi-synonyms as well as scalar and polar antonyms meet these conditions.

5.2. CASE STUDY 1: RESULTS AND ANALYSIS

5.2.1. First step: analysis of conceptual dependency

The first step of the methodology outlined in 5.1. requires the analysis of conceptual relations between frames pertaining to different levels of abstraction. The frames selected for this case study are REVEAL_SECRET and SECRECY_STATUS. A priori, these two frames have properties consistent with a relation of unidirectional conceptual dependency in the Langackerian sense of the term. They have some representational components in common but show different levels of internal complexity.

The definition of SECRECY_STATUS in FrameNet recognises only one core FE (i.e. only one semantic role which constitutes a conceptually necessary component of

the frame). This status is assigned to PHENOMENON, which stands for what is hidden from the COGNIZER:

> A Phenomenon, which may be an activity, state or object, is purposefully hidden from the awareness of a potential Cognizer. The person responsible for the hiding of the Phenomenon is not part of this frame and cannot be syntactically realized. The potential Cognizer may be realized, but is in fact rarely expressed. A period of Time during which the Phenomenon is secret may be indicated. (s.v.)

The conceptual structure represented in this definition has a high level of autonomy with respect to other events. As the above definition makes explicit, participants with an agentive semantic role are not an integral part of the frame. Hence, the structure of the frame does not include a causative layer, since the status of secrecy attributed to the PHENOMENON can be conceptualised without reference to the participants that have caused the stated situation, as the following examples from our corpus illustrate (in what follows, all the examples mentioned are from the ukWaC):[5]

(2) a. Of course there's a SECRET [recipe PHENOMENON].
b. The CLASSIFIED [document PHENOMENON], written three weeks ago, says...
c. ...an intelligence operative, [whose identity PHENOMENON] is [totally DEGREE] SECRET...
d. According to a CONFIDENTIAL [memorandum of the meeting PHENOMENON], Heath reminded Lord Widgery that...
e. He concluded that these were COVERT [missions PHENOMENON].

In comparison, the REVEAL_SECRET frame has a more complex structure. Three of the participants mentioned in the definition (SPEAKER, INFORMATION, MEDIUM) are assigned the status of core FEs in the entry, and one of them, SPEAKER, adds a layer of causation:

> A Speaker reveals Information that was previously secret to an Addressee. In some cases, the Addressee is expected to keep the Information from other parties. The Information may be damaging to the reputation of the Speaker. Instead of (or in addition to) a Speaker, a Medium may also be mentioned. Likewise, a Topic may be stated instead of Information. Some lexical units in this frame imply that the Addressee has already been confronted with the Information. (s.v.)

The structure of this frame represents a type of event which is conceptually dependent on the representation of the state of affairs depicted for the SECRECY_STATUS frame. The action of revealing the hidden information presupposes the conceptualisation of a phenomenon as being hidden, i.e. as having secrecy status.

[5] The annotation of the examples follows a conventional notation system: frame element labels are in small capitals inside square brackets, and the principal frame-evoking unit is in small capitals outside the brackets.

However, this presupposition does not hold in the opposite direction: the fact that some PHENOMENON is hidden from the awareness of a potential COGNIZER does not imply the action of a REVEALER. Secrets may or may not be disclosed. The relation between the two frames has thus the basic characteristics of A/D layering described in Section 4.2.

This observation is further reinforced by the relations of specification between FEs from the two frames. This can be illustrated through a comparison of examples (2a)–(2e) with their counterparts in (3a)–(3e). Since the action of revealing hidden information is not always performed by means of a speech act – it can also be done by submitting a file or giving access to it – the annotation of the participant acting as agent in the REVEAL_SECRET frame has been changed here from SPEAKER to REVEALER:

(3) a. We avidly await [the day TIME] [she REVEALER] REVEALS [the recipe for her brilliant pickles INFORMATION].
b. [The attorney general's office REVEALER] has LEAKED [every prosecution document INFORMATION] [to the press ADDRESSEE].
c. [Scooter Libby REVEALER] was indicted for DISCLOSING [the identity of an undercover CIA operative INFORMATION].
d. ...that [you REVEALER] will not, [directly or indirectly MANNER], DISCLOSE or permit [anyone else REVEALER] to DISCLOSE [this memorandum or its content INFORMATION] [to any other person, firm or entity ADDRESSEE]
e. ...though [the mission INFORMATION] was not DISCLOSED [to the bank officials ADDRESSEE]. [DNI REVEALER]

The core FE INFORMATION in the subordinate frame is a specification of the core FE PHENOMENON in the superordinate frame. Entities, situations, or activities that are represented as fillers of the PHENOMENON role in the frame SECRECY_STATUS are potential fillers of the INFORMATION role in the frame REVEAL_SECRET. The definitions of these FEs in the corresponding FrameNet entry can also be adduced to illustrate this point. The single core FE in the frame SECRECY_STATUS is defined as follows: "The Phenomenon is the activity, state or object that is purposefully hidden from the awareness of a potential cognizer". Its counterpart in the frame REVEAL_SECRET receives the following definition: "Information identifies the content that the Speaker reveals to the Addressee" (s.v.). The status of revealed information in the latter frame presupposes a prior conceptualisation as hidden phenomenon in the former. A similar relation holds between the non-core FEs COGNIZER and ADDRESSEE. The COGNIZER is defined as "the person who might become aware of the Phenomenon if it was not concealed" (s.v.). As long as the phenomenon remains concealed, the potential COGNIZER remains as such. It is the action of revealing the INFORMATION that turns the entity which fills out the COGNIZER slot in the superordinate frame into a filler of the

ADDRESSEE slot in the subordinate frame: "The Addressee is the person to whom the Information is revealed" (s.v.). The relation of specification holding between ADDRESSEE and COGNIZER is parallel to that holding between INFORMATION and PHENOMENON.

In sum, the analysis conducted in this section indicates that the relation between the two frames under scrutiny meets the characteristics of A/D layering. This relation of conceptual dependency is dominated by SECRECY_STATUS, which has superordinate status over REVEAL_SECRET. The structure of the latter presupposes the former, but not vice versa. The implication for the next step is that we should expect the occurrence of words evoking SECRECY_STATUS as selectional preferences of words evoking REVEAL_SECRET.

5.2.2. Second step: analysis of collocate types

The second step of the methodology was applied to three target lexical units from the subordinate frame (REVEAL_SECRET): *disclose, reveal, leak*. Following the methodological settings specified in Section 5.1., significant collocates of these three node words were extracted from the corpus. All of them met the requirement of occurring as adjectives in object noun phrases of the three verbs. The three lists of collocates, arranged in order of decreasing score, are displayed in the tables below. Table 1 and Table 2 show the top 50 collocates. In the case of *leak*, there was no need to establish a limit on the maximum number of items, because the number of collocates with positive association scores was smaller.

The results indicate the presence of evokers of the superordinate frame in the three lists. In Table 1, there are ten adjectival collocates capable of evoking the frame SECRECY_STATUS. These are the following: *classified, confidential, hidden, intimate, personal, private, privileged, secret, sensitive*, and *unpublished*. In Table 2, we find six adjectives capable of evoking the same frame: *confidential*, *hidden, inner, innermost, intimate*, and *secret*. In Table 3, there are five adjectives from this set: *classified, confidential, internal, secret, sensitive*. In Tables 2 and 3, we can also observe the presence of other semantic types of collocates. This can be related to the polysemy of the node. The verbs *reveal* and *leak* are attributed to more than one frame in FrameNet. In addition to SECRECY_STATUS, *reveal* can also evoke the frame EVIDENCE, and *leak* can also evoke the frame FLUIDIC_MOTION. This semantic variation is reflected in the collocate inventory. In Table 2 we find several adjectival collocates describing an 'intense emotional and cognitive impact on an experiencer' (e.g. *astonishing, disturbing, fascinating, shocking, startling, staggering, striking, stunning*). These form part of some typical linguistic realisations of the EVIDENCE frame (e.g. *...has revealed disturbing cases of self-harm; ...has revealed striking variations in the training...; ...has revealed fascinating differences...*). Similarly,

Table 1. Top 50 collocates of *disclose* (word class: adjective; syntactic context: premodifer of object noun).

collocate (lemma)	logDice	collocate (lemma)	logDice
confidential	7.490	*hidden*	2.825
personal	6.034	*privileged*	2.823
prima facie	5.846	*reasonable*	2.790
classified	5.277	*actual*	2.652
sensitive	5.151	*previous*	2.545
exempt	5.113	*sexual*	2.446
unused	4.592	*widespread*	2.400
exact	4.300	*beneficial*	2.391
contingent	4.136	*financial*	2.343
pecuniary	4.121	*unexpected*	2.296
distressing	3.999	*specific*	2.229
proprietary	3.850	*adverse*	2.149
relevant	3.784	*true*	2.143
procedural	3.500	*medical*	2.120
criminal	3.463	*divine*	2.070
secret	3.443	*sufficient*	1.994
malignant	3.397	*significant*	1.832
intimate	3.375	*gross*	1.832
pertinent	3.201	*partial*	1.796
precise	3.200	*domestic*	1.706
certain	3.130	*private*	1.623
geographical	3.088	*serious*	1.510
genetic	3.055	*inner*	1.440
unpublished	2.939	*statistical*	1.378
material	2.862	*full*	1.321

in Table 3 we find some adjectives related to the notion of 'danger' (*dangerous, deadly, harmful*). These are associated with the activation of the FLUIDIC_MOTION frame (*...leaking dangerous gas, ...leak deadly carbon monoxide, ...leak harmful chemicals,* etc.). The fact that adjectives related to 'secrecy' constitute the only dominant semantic type in Table 1 may be related to the greater specialization of

Table 2. Top 50 collocates of *reveal* (word class: adjective; syntactic context: premodifer of object noun).

collocate (lemma)	logDice	collocate (lemma)	logDice
hidden	7.422	*considerable*	4.799
true	6.402	*distinct*	4.794
secret	6.296	*dramatic*	4.791
shocking	6.114	*exact*	4.751
startling	5.884	*fundamental*	4.707
fascinating	5.772	*substantial*	4.700
significant	5.743	*profound*	4.664
unexpected	5.658	*ambitious*	4.616
astonishing	5.562	*innermost*	4.511
inner	5.547	*serious*	4.460
widespread	5.527	*extensive*	4.395
interesting	5.428	*huge*	4.372
surprising	5.418	*complex*	4.343
extraordinary	5.408	*massive*	4.321
worrying	5.331	*overwhelming*	4.316
disturbing	5.176	*confidential*	4.306
alarming	5.108	*divine*	4.291
striking	5.089	*dark*	4.274
marked	5.071	*enormous*	4.261
intimate	4.948	*structural*	4.225
deep	4.947	*stunning*	4.087
remarkable	4.943	*strong*	4.073
unsuspected	4.908	*sensitive*	4.065
insight	4.901	*appalling*	4.065
subtle	4.825	*devastating*	4.037

disclose in relation to the SECRECY_STATUS frame. In fact, in the FrameNet database, *disclose* is exclusively attributed to this frame.

In the current FrameNet release, the relation between the adjectives from Tables 1-3 and the SECRECY_STATUS frame is only partially documented. The adjectives

Table 3. Collocates of *leak* (word class: adjective; syntactic context: premodifer of object noun).

collocate (lemma)	logDice	collocate (lemma)	logDice
classified	6.297	*sensitive*	2.809
confidential	4.538	*secret*	2.329
hydraulic	4.012	*chemical*	2.315
harmful	3.068	*false*	1.680
internal	2.963	*dangerous*	1.064
deadly	2.843	*official*	0.543

confidential and *secret* are provided as lexical units in the entry for SECRECY_STATUS, but the other eleven adjectives (*classified, hidden, inner, innermost, internal, intimate, personal, private, privileged, sensitive*, and *unpublished*) are not. There is evidence to argue that they can be added to the list, since they comply with the general FrameNet criteria for being grouped into the same frame. The reasons for this are explained below in some detail.

The first reason concerns the similarity of semantic type. The eleven adjectives under scrutiny belong to a similar semantic type as other lexical units considered as evokers of the SECRECY_STATUS frame (i.e. *confidential* and *secret*). The lexicographic evidence for this is twofold. It can be found both in the synonym sets provided by the thesaurus and in the meaning definitions provided by general-purpose dictionaries. In the OTE, all these adjectives form part of the synonym set for at least one of the lexical units described in the FrameNet entry for this frame. This is shown in Table 4. Each column is occupied by one of the two *reference collocates* (i.e. *confidential* and *secret*). These are the lexical units that occur both in the collocate inventory (Tables 1-3) and in the FrameNet entry for the frame under scrutiny. Each row corresponds to one of the *target collocates*, i.e. those lexical units from the collocate inventory which can evoke the same frame as the reference units but are not documented in the FrameNet entry. Cells marked with "R" indicate those cases where the target collocates are found in the synonym entries for the reference collocates; cells marked with "T" correspond to cases where the reference collocates are found in the synonym entries for the target collocates. The tables include only those collocates which have at least one match in one of the synonym entries considered.

Some further refinement was necessary in order to deal with the problem of polysemy. This required discriminating between senses of the adjectives in the

collocations and in the thesaurus entries. For instance, *deep* was not included in Table 4, even though it occurs in one of the collocate lists and has *secret* as a synonym of one of its senses in the OTE. The reason for this is that its occurrences as a collocate of *reveal* are not specialised in the 'unknown' sense of this adjective. In collocation with *reveal*, *deep* is also frequently used as an intensifier (e.g. *Brown recently revealed his deep disquiet at the government deal*; *...it reveals the deep hatred levelled at those who...*; *...has revealed a deep level of disquiet with...*). *Dark* was also excluded for similar reasons. Its co-occurrences with *reveal* are not specialised in the SECRECY_STATUS frame (e.g. *...and revealed a small dark chamber just inside; ...a strand of his hair was sticking out, ...revealing a dark raven color*).

The information shown in Table 4 suggests that the eleven target collocates (*classified, hidden, inner, innermost, internal, intimate, personal, private, privileged, sensitive, unpublished*) can be grouped into the same semantic type as other units which evoke the SECRECY_STATUS frame. Their presence in this table indicates that, for each of them, it is possible to find at least one relation of synonymy with a reference collocate.

A second type of thesaurus information which is also relevant for our analysis concerns oppositeness. At this point, it is important to recall that polar and scalar antonyms are not split into separate frames in FrameNet, because they do not profile different participants (Ruppenhofer et al. 2016). In relation to our case

Table 4. Pairings of target and reference collocates in thesaurus entries.

		Reference collocates	
		confidential	*secret*
Target collocates	*classified*	R	R
	hidden		R/T
	inner	T	T
	innermost		T
	internal		T
	intimate	R/T	T
	personal	R/T	
	private	T	R/T
	privileged	R/T	T
	sensitive	R	
	unpublished	R	R

study, this is especially relevant for the analysis of *private*. The adjectives *open* and *public* are registered as target lexical units from the SECRECY_STATUS frame, and *private* is an antonym of these two adjectives. This relation is in fact recorded in the OTE, where *private* appears as an opposite in the entries for *public* and *open*, and conversely, *public* and *open* are included as opposites in the entry for *private*. This can be interpreted as further evidence for the treatment of *private* as a unit capable of evoking the SECRECY_STATUS frame.

The semantic affinity between the target collocates and the reference collocates is further reinforced by their definitions. Tables 5 and 6 show excerpts from entries in three different dictionaries (in the case of polysemous items, only the most relevant senses for the frame under scrutiny have been cited). The relation with the reference collocates is straightforward in the definitions of two of the target collocates, namely, *classified* and *sensitive*. As can be observed, their definiens in the three dictionaries includes the word *secret* itself. For another group of target collocates (*inner, innermost, private, privileged*), the word *secret* is used in only one of the three definitions. Finally, there is another group in whose definitions the word *secret* is not present at all. This is the case of *hidden, internal, intimate, personal*, and *unpublished*. However, there are other signs of their affinity with the meaning of the reference collocates, which can be captured through a closer examination of their definitions.

A common thread running through the definitions of the two reference collocates (Table 5) is the idea of a 'restricted access to knowledge'. This notion is also present, with various nuances, in the meaning potential of the target collocates (Table 6). Thus, in some definitions we find paraphrases such as "not known or available to most people" (*personal*, MED); "not to be revealed to others" (*private*, ODE); "not available for the public to read" (*unpublished*, MED); "most people do not know about it or understand it" (*hidden*, MED); "not allowed to be made public by law" (*privileged*, LDOCE); etc. Admittedly, this notion of "restricted access to knowledge" is less obvious in some of the definitions. *Internal* is a case in point. However, this is mainly due to the fact that, in some cases, the dictionary cannot cover the entire range of arguments with which the word is combined. Thus, in the case of *internal*, we can observe that some of the collocational patterns involve combinations with nouns denoting 'information or documents', e.g. *memo, email, document, report*. This is in fact the collocational pattern of *internal* that is habitually merged with *leak* (e.g. *...recently leaked an internal report; ...had leaked an internal memo; ...leaking an internal document*). Used in combination with this type of argument, the specific sense of *internal* is very similar to that of *classified*, since it restricts the domain within which a document is allowed to be circulated.

Table 5. Definitions of reference collocates from the SECRECY_STATUS frame.

Lemma	Word sense definitions		
	ODE	**LDOCE**	**MED**
confidential	intended to be kept secret	spoken or written in secret and intended to be kept secret	confidential documents or information must be kept secret
secret	not known or seen or not meant to be known or seen by others [...] not meant to be known as such by others	known about by only a few people and kept hidden from others [...] secret feelings, worries or actions are ones that you do not want other people to know about	deliberately not told to other people or kept hidden from other people

Table 6. Definitions of senses of target collocates related to the SECRECY_STATUS frame.

Lemma	Word sense definitions		
	ODE	**LDOCE**	**MED**
classified	(of information or documents) designated as officially secret and accessible only to authorized people	classified information, documents etc are ones which the government has ordered to be kept secret	classified information is officially secret and allowed to be known by only a few people connected with the government or armed forces
hidden	kept out of sight; concealed	not easy to notice or realize	if something is hidden, most people do not know about it or understand it
inner	(of thoughts or feelings) private and not expressed or discernible; denoting a concealed or unacknowledged part of a person's personality	relating to things which happen or exist but are not easy to see	private, personal, or secret

innermost	(of thoughts or feelings) most private and deeply felt	your innermost feelings, desires etc are your most personal and secret ones	your innermost thoughts and feelings are the ones that are most personal and private
internal	existing or occurring within an organization	within a company or organization rather than outside it	existing or happening within an organization or institution
intimate	private and personal	private and friendly so that you feel comfortable	relating to very private or personal things
personal	of or concerning one's private life, relationships, and emotions rather than one's career or public life	relating to the private areas of your life	private and not known or available to most people
private	(of a conversation, activity, or gathering) involving only a particular person or group, and often dealing with matters that are not to be disclosed to others [...] (of thoughts and feelings) not to be revealed to others	a private meeting, conversation etc involves only two people or a small number of people, and is not for other people to know about [...] private feelings, information, or opinions are personal or secret and not for other people to know about	understood by only a few people, not by everyone
privileged	(of information) legally protected from being made public	privileged information is private and is not allowed to be made public by law	privileged information is secret and does not have to be discussed
sensitive	kept secret or with restrictions on disclosure to avoid endangering security	a situation or subject that is sensitive needs to be dealt with very carefully, because it is secret or because it may offend people	needing to be kept secret
unpublished	(of a piece of writing or music) not issued in print for public sale or consumption	unpublished writing, information etc has never been published	not available for the public to read

In sum, all of the target collocates can be used to describe some kind of difficulty in accessing a particular aspect of information or knowledge. The object of that knowledge corresponds to the PHENOMENON element in the SECRECY_STATUS frame, as well as to the role INFORMATION in the REVEAL_SECRET frame; the person whose access to that knowledge is restricted is represented by the COGNIZER in the SECRECY_STATUS frame, and by the ADDRESSEE in the REVEAL_SECRET frame. This configuration of semantic roles can be applied to at least some patterns of use of all the target collocates. This leads us to the next reason for same-frame grouping, which is *near-paraphrasability*. As defined in FrameNet, this criterion refers to felicitous substitutability in syntactic environments characterised by the similar configurations of frame elements. Examples (4)-(5) illustrate this possibility with a sample of target collocates. The annotation of these examples integrates descriptive labels of elements from the subordinate and the superordinate frames (to distinguish them from other labels, the evokers of frames from different levels are marked with a subscript). Thus, in example (4), the element PHENOMENON and the adjectives that evoke it (i.e. the reference collocates *secret* and *confidential*) are embedded into a core element (INFORMATION) evoked by the principal bearer of the subordinate frame (in these examples, the verb *disclose*). Additionally, two elements of these frames (ADDRESEE and COGNIZER) converge into a single constituent. This complex configuration of semantic roles can be accompanied by other elements that belong exclusively to one of the frames combined. In these examples, the element REVEALER is exclusively activated by the subordinate frame. Crucially, examples (5) show a similar configuration of semantic roles – and also a similar mapping onto valency patterns – with the target collocates acting as evokers of the superordinate frame.

(4) a. ...after [he REVEALER] DISCLOSED$_{SUB}$ [SECRET$_{SUP}$ [MI5 documents PHENOMENON] INFORMATION] [to a British newspaper ADDRESSEE/COGNIZER].
b. [We REVEALER] had DISCLOSED$_{SUB}$ [some CONFIDENTIAL$_{SUP}$ [sales data PHENOMENON] INFORMATION] [to one of our suppliers ADDRESSEE/COGNIZER] on the clear understanding that it remained confidential.

(5) a. ...her first husband (...) demanded she paid him the staggering sum or [he REVEALER] would REVEAL$_{SUB}$ [INTIMATE$_{SUP}$ [details of their relationship PHENOMENON] INFORMATION] [to the media ADDRESSEE/COGNIZER].
b. [Who ADDRESSEE/COGNIZER] do [we REVEALER] DISCLOSE$_{SUB}$ [your PERSONAL$_{SUP}$ [details PHENOMENON] INFORMATION] to?
c. When [a person REVEALER] DISCLOSES$_{SUB}$ [PRIVATE$_{SUP}$ [data PHENOMENON] INFORMATION] [to an organization ADDRESSEE/COGNIZER], accuracy is obviously an ethical issue.
d. Sometimes [the taxing officer REVEALER] will have to DISCLOSE$_{SUB}$ [PRIVILEGED$_{SUP}$ [documents PHENOMENON] INFORMATION] [to the other side ADDRESSEE/COGNIZER].
e. [DPTAC and the DRC REVEALER] will not DISCLOSE$_{SUB}$ [UNPUBLISHED$_{SUP}$ [information

> PHENOMENON] INFORMATION] [to third parties ADDRESSEE/COGNIZER] without the permission of the organisation providing the information.

5.3. CASE STUDY 2: RESULTS AND ANALYSIS

5.3.1. First step: analysis of conceptual dependency

This second case study is focused on relations between the dependent frame COMMITTING_CRIME and two other frames, LEGALITY and MORALITY_EVALUATION. The three frames share a substantial part of their conceptual structure, but they are located on different levels of specificity: the COMMITTING_CRIME presupposes the LEGALITY frame as background, and in turn, the LEGALITY frame presupposes the MORALITY_EVALUATION frame. This is coded in FrameNet in the form of successive "Using" relations connecting these entries. "Using" relations are those in which "a particular frame makes reference in a very general kind of way to the structure of a more abstract, schematic frame" (Ruppenhofer et al. 2016: 83). In the FrameNet database, MORALITY_EVALUATION is described as being "used by" LEGALITY, and LEGALITY is described as being "used by" COMMITTING_CRIME.

Additionally, the relationship between the first frame and the other two shows the characteristics of conceptual dependency explained in Section 4.2. Observe the definition of COMMITTING_CRIME in FrameNet:

> A Perpetrator (generally intentionally) commits a Crime, i.e. does something not permitted by the laws of society. (s.v.)

The conceptual structure represented in this definition includes an agentive participant (PERPETRATOR) which is attributed coreness status. This participant has no equivalent element in the structure of the other two frames. Their definitions are quoted below:

> Words in this frame [LEGALITY] describe the status of an Action with respect to a Code of laws or rules. An Object may also be in violation or compliance of the Code by virtue of its existence, location or possession. (s.v.)
> In this frame [MORALITY_EVALUATION] an Evaluee is described by a (usually implicit) Judge with respect to the morality or rightness of his or her Behavior. (s.v.)

With respect to these other two frames – both of which are situated in the same chain of "Using" relations – the COMMITTING_CRIME frame adds a layer of causation, represented by the agentive role of the PERPETRATOR. In this sense, COMMITTING_CRIME is a subordinate frame relative to LEGALITY and MORALITY_EVALUATION. Given the relationship of presupposition between these two frames (and their superordinate status with respect to COMMITTING_CRIME), we can expect to find

their frame evokers among the selectional preferences of items which evoke the COMMITTING_CRIME frame. This is explored in the next section.

5.3.2. Second step: analysis of collocate types

The entry for COMMITTING_CRIME in FrameNet records four lexical units, of which two are verbs (*commit* and *perpetrate*). The two verbs have been entered here as nodes for collocate extraction, following the methodological settings specified in Section 5.1. The resulting list of adjectival collocates is displayed in Tables 7-8. Table 7 shows the top 50 collocates obtained with *commit* as node. In the case of *perpetrate*, there was no need to establish a limit on the maximum number of items, because the number of collocates with positive association scores was smaller.

The results in Table 7-8 show a predominance of adjectives expressing a negative evaluation. These extend along a cline ranging from more specific to more schematic meanings. On the more specific pole of the cline are those adjectives that are capable of evoking the LEGALITY frame. These are the adjectives *criminal, delinquent, fraudulent, illegal, indictable, unlawful* and *wrongful*. Table 7 contains all these adjectives; Table 8 contains only two of them (*criminal* and *illegal*). Another set of adjectives is formed by words which convey or are typically associated with a negative judgment, but which do not include the notion of 'legality' as an essential component of their meaning. The list is extensive. Table 7 contains 25 of them: *abominable, appalling, atrocious, awful, brutal, despicable, detestable, dreadful, egregious, evil, grave, grievous, gross, gruesome, heinous, horrendous, horrific, immoral, indecent, inhumane, serious, terrible, unforgivable, unpardonable, unspeakable*. Table 8 contains 11 of these adjectives: *appalling, cruel, dreadful, evil, gross, heinous, horrible, horrific, inhuman, malicious, unspeakable*.

As in the previous case study, we can observe that the capacity of the collocates for evoking the superordinate frames is only partially documented in the corresponding FrameNet entries. The adjectives *criminal, illegal, unlawful* and *wrongful* are provided as lexical units in the entry for LEGALITY, but there are three other adjectives in Tables 7-8 (*delinquent, fraudulent, indictable*) which are not recorded in this entry and show a strong capacity for evoking the same frame. As in Section 5.2.2., I will use the terms *reference collocates* and *target collocates* to distinguish these two groups. The semantic affinity between the two can be identified by means of the same criteria applied above, which include the analysis of information from a thesaurus, from meaning definitions in general-purpose dictionaries, and from the analysis of example concordances. Concerning the first

Table 7. Top 50 collocates of commit (word class: adjective; syntactic context: premodifer of object noun).

collocate (lemma)	logDice	collocate (lemma)	logDice
criminal	7.971	*brutal*	5.292
heinous	7.723	*horrific*	5.261
violent	7.208	*horrible*	5.260
serious	6.399	*dreadful*	5.116
alleged	6.330	*unforgivable*	5.074
terrible	6.295	*acquisitive*	4.953
detestable	6.276	*minor*	4.829
unlawful	6.177	*evil*	4.760
indecent	6.095	*penal*	4.733
indictable	5.978	*substantial*	4.651
horrendous	5.949	*murderous*	4.642
unspeakable	5.911	*immoral*	4.492
atrocious	5.820	*non-violent*	4.480
gross	5.784	*fraudulent*	4.424
sexual	5.761	*unprovoked*	4.403
grievous	5.738	*abominable*	4.384
anti-social	5.729	*inhumane*	4.362
imprisonable	5.686	*delinquent*	4.349
unpardonable	5.670	*despicable*	4.345
wrongful	5.632	*awful*	4.276
mortal	5.575	*hostile*	4.244
grave	5.569	*egregious*	4.208
unnatural	5.525	*disciplinary*	4.198
illegal	5.367	*gruesome*	4.193
appalling	5.299	*non-political*	4.177

criterion, observe that for each of the three target collocates, the OTE provides at least one relation of synonymy with a reference collocate (Table 9). The definitions of the target collocates also attest to their relation with the semantic type of the reference collocates. In these definitions, words such as *illegal*, *criminal*, *crime*, and *offence* establish the 'illegal' status of one of the arguments of the adjectives (Table 10). Finally, examples (6)-(7) illustrate the possibility of using target and

Table 8. Top 50 collocates of *perpetrate* (word class: adjective; syntactic context: premodifer of object noun).

collocate (lemma)	logDice	collocate (lemma)	logDice
unspeakable	5.938	*malicious*	3.569
senseless	5.222	*cruel*	3.270
heinous	5.142	*anti-social*	2.991
inhuman	5.125	*gross*	2.971
untold	4.638	*sexual*	2.930
brutal	4.053	*criminal*	2.183
dreadful	3.942	*domestic*	1.922
evil	3.924	*massive*	1.641
violent	3.919	*illegal*	1.622
horrific	3.780	*false*	0.929
appalling	3.701	*numerous*	0.887
horrible	3.639	*moral*	0.680

reference collocates in environments with a similar configuration of semantic roles and with a similar valency pattern (as with examples (4)-(5), the annotation shows embedding of superordinate frame elements within the structure of the subordinate frame).

(6) a. [You PERPETRATOR] have COMMITTED$_{SUB}$ [an ILLEGAL$_{SUP}$ [operation ACTION] CRIME].
b. ...protesters [who PERPETRATOR] had [sometimes TIME] COMMITTED$_{SUB}$ [CRIMINAL$_{SUP}$ [acts ACTION] CRIME].

(7) a. ...shows that [boys PERPETRATOR] are still more likely to COMMIT$_{SUB}$ [DELINQUENT$_{SUP}$ [acts ACTION] CRIME] than girls.
b. ...too easy for [a criminal PERPETRATOR] to COMMIT$_{SUB}$ [FRAUDULENT$_{SUP}$ [transactions ACTION] CRIME].
c. Should [a Party member PERPETRATOR] COMMIT$_{SUB}$ [an INDICTABLE$_{SUP}$ [offence ACTION] CRIME], he shall be expelled from the Party.

In general, both verbs (*commit* and *perpetrate*) show a strong preference for adjectival collocates with a negative evaluative meaning. Some of them (e.g. *terrible, horrible, horrific, serious*, *grave*, etc.) have highly schematic meanings and can be applied to a broad variety of argument classes, not all of which are endowed with morally assessable qualities (e.g. *horrific nightmares, terrible disaster, horrible wounds, gruesome fate*, *serious damage, grave concern,* etc.).

Table 9. Pairings of target and reference collocates in thesaurus entries.

		Reference collocates			
		criminal	*illegal*	*unlawful*	*wrongful*
Target collocates	*delinquent*	R/T		R	
	fraudulent	R/T	*T*	R/T	
	indictable	R		R	

Table 10. Definitions of senses of target collocates related to the LEGALITY frame.

Lemma	**Word sense definitions**		
	ODE	**LDOCE**	**MED**
delinquent	(typically of a young person) tending to commit crime, particular minor crime	behaving in a way that is illegal or that society does not approve of	behaving in a way that is criminal or antisocial
fraudulent	obtained, done by, or involving deception, especially criminal deception	intended to deceive people in an illegal way, in order to gain money, power etc.	made with the intention of tricking someone, especially illegally
indictable	(of an offence) rendering the person who commits it liable to be charged with a serious crime that warrants a trial by jury	an indictable offence is one for which you can be indicted	an indictable offence is one for which you can be officially accused and brought to a court for trial

Following the FrameNet taxonomy of semantic types, these adjectives could be marked as bearing NEGATIVE_JUDGMENT. However, it should be noted that this aspect of meaning is applied to lexical units across semantic frames (Ruppenhofer et al. 2016). In its syntagmatic dimension, this semantic property corresponds to what Bednarek (2008) describes as a "NEG collocation" pattern, which is a specific manifestation of the more general phenomenon of *semantic preference/prosody*. However, other collocates of *commit* and *perpetrate* are more specific and show a

Table 11. Pairings of target and reference collocates in thesaurus entries.

		Reference collocates		
		evil	*heinous*	*immoral*
Target collocates	*abominable*		R	
	cruel	T	T	
	despicable	R	R/T	
	indecent			*R*
	inhuman		*T*	
	inhumane		*T*	
	malicious	*R*		*R*
	unforgivable		*R*	
	unpardonable		*R*	

potential for evoking the MORALITY_EVALUATION frame. Three of them (*evil, heinous, immoral*) are in fact attributed to this frame in FrameNet. Other collocates from Tables 7-8 can also be added to this set, since the negative judgement they express is systematically applied to the moral or ethical qualities of particular actions, or of the human individuals or institutions responsible for such actions. Good examples of this are *abominable, cruel, despicable, indecent, inhuman, inhumane, malicious, unforgivable,* and *unpardonable.* These adjectives meet the condition of near-paraphrasibility in relation to *evil, heinous* and *immoral*, as examples (8)-(9) illustrate. Moreover, as Table 11 shows, they can be classified as synonyms of one or more reference collocates (*evil, heinous, immoral*).

(8) a. We want to forget and forgive those [who PERPETRATOR] PERPETRATED$_{SUB}$ [this HEINOUS$_{SUP}$ [crime ACTION] CRIME].
b. Those [who PERPETRATOR] PERPETRATED$_{SUB}$ [this EVIL$_{SUP}$ [attack ACTION] CRIME] must be brought to justice.

(9) a. [These men PERPETRATOR] COMMITTED$_{SUB}$ [their DESPICABLE$_{SUP}$ [crimes ACTION] CRIME] [by targeting the most vulnerable members of society MANNER].
b. ...its him [who PERPETRATOR]'s COMMITTED$_{SUB}$ [those INHUMAN$_{SUP}$ [war crimes ACTION] CRIME].

5.4. DISCUSSION

To frame the significance of the results analysed in section 5.2., it will be useful to recall the standard syntactic types of collocational patterns (section 2). These include combinations of a predicate and the head of one of its arguments: V+N, Adj+N, Adv+V, Adv+Adj, and N+Prp+N (Corpas Pastor 1996; Hausmann 1998; Bosque 2001a, 2004a, 2004b; Martin 2008; Tutin 2008). This model has been extended to the analysis of *ternary* (o *tripartite*) collocations. Ternary combinations with the structure V+Adj+N are decomposed into two collocational pairs, V+N and Adj+N, which are constituted independently of the co-occurrence of the verb and the adjective in the extended pattern (Hausmann 2004; Alonso Ramos and Wanner 2007). However, the patterns analysed in the previous section do not fit into this model, because they show a semantic link between a predicate and a non-head constituent of an argument.

Selectional relations between verbs and modifiers of nouns have not been frequently described in the literature on collocation. Almela-Sánchez (2011) and Almela-Sánchez y Cantos-Gómez (2018) focus on corpus-based techniques for capturing patterns of lexical association between verbs and adjectives in complex collocations with nouns. Bosque (2001b) and Koike (2001) offer an analytical framework for cases where the selectional constraint operates between the verb and the modifier of the noun, as in Sp. *eludir una empresa peligrosa* ('to avert a dangerous undertaking'), *enfrentarse a situaciones peligrosas* ('to face dangerous situations'), *pasar por circunstancias críticas* ('to go through critical circumstances'), or *desencadenar una ola de violencia* ('to unleash a wave of violence'), among others. Although these authors apply different theoretical frameworks – Bosque subsumes collocation within the Chomskyan notion of *s-selection*, while Koike is in line with the Hausmann-Mel'čuk phraseological approach – their analysis of ternary combinations converge on a similar idea: the selectional link between the verb and the modifier of the noun is concomitant with nouns that are semantically downgraded and that fail to intervene in the collocational link with the verb. Koike (2001) observes that in the aforementioned examples, the head noun (*empresa, situación, circunstancia, ola*) has a *semantically neutral quality*: "un rasgo semánticamente neutro" (Koike 2001: 160). Bosque (2001b) agrees with this statement and offers a similar formulation of the same phenomenon: in these examples, the abstract nouns "provide 'instances' of the quality or state of affairs denoted by the adjective" (Bosque, 2001b: 33). Bosque (2001b) uses the term *light noun* – by analogy to *light verb* – to characterise the behaviour of this type of items.

However, the characteristics of collocations with light nouns do not apply to the patterns analysed in section 5.2. The verb-adjective collocations described

there are compatible with nouns that intervene in the collocation with the verb (*disclose a memorandum, disclose data, leak documents*, *reveal information*, etc.; *commit a crime, perpetrate an attack*, etc.) and whose semantic content is not reducible to the schematicity of a light noun. Rather than serving merely as a means to support the semantic content of the modifier (*secret, confidential, privileged, sensitive*, etc.), nouns such as *memorandum, data, documents*, *information*, etc., constitute themselves another semantic set, organised around the notion of INFORMATION, which reflects a further aspect of the selectional preferences of the verb. In ternary combinations such as *disclose a confidential memorandum, disclose privileged documents, disclose sensitive data, leak classified documents, reveal secret information*, *commit despicable crimes, perpetrate an evil attack, perpetrate heinous crimes*, etc., both the adjective and the noun provide a lexical realisation of semantic preferences of the verb. Rather than cases of collocations with light nouns, these examples seem to reflect a phenomenon of *stratified collocation*, whereby the predicate is involved simultaneously in a collocational link with the head and with a non-head constituent of the argument phrase.

There is some *prima facie* evidence that stratified collocational patterns such as those pointed out here may not be isolated cases. It is possible to allude to examples which, a priori, seem to exhibit similar properties. Koike (2001) mentions some examples which are ambiguous between verb-noun and verb-adjective links. One of them is Sp. *satisfacer al estómago más exigente* ('to satisfy the most demanding guest/consumer', where *estómago* (*stomach*) is interpreted metonymically). Koike (2001) concedes that in this example it is difficult to determine whether the lexical bond is established between the verb and the noun, or between the verb and the modifier of the noun. Arguably, what this ambiguity reflects is simultaneity: *satisfy* collocates both with nouns referring to 'somebody who receives a service' (*client, customer, user*, *consumer*, *patient*, etc.) and with nouns and adjectives that describe the 'setting of a standard or objective' (*demand*, *requirement, condition, constraint, need, craving, request, hunger, appetite*...; *demanding, insatiable, hungry, stringent*...). The relationship between the two sets of preferences fits readily into A/D layering, because the setting of a standard is a conceptually more basic structure than the process of successfully meeting the standard (the latter presupposes the former, but not the other way round). Again, in these examples, the noun and the modifier of the noun realise different sets of selectional preferences of the same valency bearer. Further corpus research will determine whether these and other examples can indeed be interpreted as additional empirical evidence for A/D layering in the realm of collocations.

6. CONCLUSIONS, LIMITATIONS, AND FURTHER RESEARCH

This article is conceived as a contribution to the development of a frame-based, valency-oriented approach to collocation. I have argued that the Langackerian notion of A/D layering and its application in the field of semantic frame description provides an efficient explanatory framework for a type of collocational pattern, which may be called *stratified collocation*, that does not fit into the standard syntactic typologies of collocation. The theoretical basis for the framework is made of correspondences between the structure of canonical valency relations and the descriptive categories of Fillmorean frame semantics. The articulation of these relations reduces to three main ideas: (*i*) the configuration of canonical valency relations in terms of autonomy-dependency asymmetries; (*ii*) the deployment of autonomy-dependency relations along two different axes (predicate-argument dependency and event dependency); and (*iii*) the default mapping of frame-evokers and FE fillers to valency bearers and valency fillers, respectively. The combination of these three claims allows us to predict the presence of evokers of a superordinate frame among selectional preferences of a subordinate frame, as well as the syntactic realisation of such selectional preferences through embedded predicate-argument structures, such as V((Adj(N))).

The empirical evidence analysed here has been focused on two specific case studies. I have also mentioned additional examples that seem to exhibit similar characteristics and that are worth exploring in forthcoming case studies. This will help us determine the extent to which dependency relations between semantic frames with different levels of complexity may be responsible for deriving productive patterns of stratified collocation.

In addition to exploring further empirical evidence, future research into stratified collocation should also be geared to clarify the configuration of relations between the different sets of selectional preferences observed in this type of patterns. In particular, one issue which deserves special attention is the role played by both predicates – verb and adjective – in establishing the selectional preferences imposed on the argument head (the noun). It remains to be ascertained whether the preferred semantic types for the noun slot result from coordinating selectional preferences of verb and adjective or from the dominant role of one of these two items. Clarifying this issue will be useful for obtaining a more accurate knowledge of how semantic and syntactic layers of valency patterning interact in frame-to-frame relations.

REFERENCES

Akita, K. 2012. "Toward a frame-semantic definition of sound-symbolic words: A collocational analysis of Japanese mimetics". *Cognitive Linguistics* 23 (1): 67-90.

Almela-Sánchez, M. 2011. "The case for verb-adjective collocations: Corpus-based analysis and lexicographical treatment". *Revista de Lingüística y Lenguas Aplicadas* 6: 39-51.

Almela-Sánchez, M. and P. Cantos-Gómez. 2018. "Towards a model of co-collocation analysis: Theory, methodology, and preliminary results". *Corpora and Lexis*. Eds. S. Hoffmann, A. Sand, S. Arndt-Lappe and L. M. Dillmann. Leiden and Boston: Brill. 209-239.

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

Alonso Ramos, M. 2017. "Can collocations be deduced? A lexically driven analysis from the perspective of language production". *Collocations and other Lexical Combinations in Spanish: Theoretical, Lexicographical and Applied Perspectives*. Eds. S. Torner Castells and E. Bernal. London and New York: Routledge. 21-40.

Alonso Ramos, M., Rambow, O. and L. Wanner. 2008. "Using semantically annotated corpora to build collocation resources". *Proceedings of the 6th International Conference on Language Resources and Evaluation, LREC 2008*. Eds. N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis and D. Tapias. 1154-1158. <http://www.lrec-conf.org/proceedings/lrec2008/index.html>. (Accessed 15 March 2018).

Alonso Ramos, M. and L. Wanner. 2007. "Collocation chains: How to deal with them?". *Wiener Slawistischer Almanach. Sonderband* 69: 11-20.

Apresjan, J. D. and M. J. Glovinskaja. 2007. "Two Projects: English ECD and Russian Production Dictionary". *Wiener Slawistischer Almanach. Sonderband* 69: 31-46.

Apresjan, J. D. 2009. "The theory of lexical functions: An update". *Proceedings of the Fourth International Conference on Meaning-Text Theory*. Eds. D. Beck, K. Gerdes, J. Milićević, Jasmina and A. Polguère. Observatoire de Linguistique Sense-Texte. 1-14. <http://olst.ling.umontreal.ca/pdf/ProceedingsMTT09.pdf>. (Accessed 8 November 2016).

Baker, C. F. 2009. "La sémantique des cadres et le projet FRAMENET: une approche différente de la notion de valence". *Langages* 176 (4): 32-49.

Bednarek, M. 2008. "Semantic preference and semantic prosody re-examined". *Corpus Linguistics and Linguistic Theory* 4 (2): 119-139.

Bartsch, S. and S. Evert. 2014. "Towards a Firthian notion of collocation". *OPAL – Online Publizierte Arbeiten zur Linguistik* 2/2014: 48-61. <http://pub.ids-mannheim.de/laufend/opal/pdf/opal2014-2.pdf>. (Accessed 23 March 2018).

Boas, F. 2001. "Frame semantics as a framework for describing polysemy and syntactic structures of English and German motion verbs in contrastive computational

lexicography". *Proceedings of the Corpus Linguistics 2001 Conference*. Eds. P. Rayson, A. Wilson, T. McEnery, A. Hardie, S. Khoja. Lancaster University. 64-73. <http://ucrel.lancs.ac.uk/publications/CL2003/CL2001%20conference/papers/boas.pdf>. (Accessed 2 September 2018).

Bosque, I. 2001a. "Sobre el concepto de 'colocación' y sus límites". *Lingüística Española Actual* 23 (1): 9-40.

Bosque, I. 2001b. "On the weight of light predicates". *Features and Interfaces in Romance: Essays in Honor of Heles Contreras*. Eds. J. Herschensohn, E. Mallén and K. Zagona. Amsterdam: John Benjamins. 23-38.

Bosque, I. 2004a (dir.). *REDES. Diccionario combinatorio del español contemporáneo*. Madrid: SM.

Bosque, I. 2004b. "La direccionalidad en los diccionarios combinatorios y el problema de la selección léxica". *Lingüística teòrica: anàlisi i perspectives I*. Ed. T. Cabré Monné. Bellaterra: Universitat Autónoma de Barcelona. 13-58.

Bosque, I. 2017. "On the conceptual bases of collocations: restricted adverbs and lexical selection". *Collocations and other Lexical Combinations in Spanish: Theoretical, Lexicographical and Applied Perspectives*. Eds. S. Torner Castells and E. Bernal. London and New York: Routledge.

Buendía Castro, M. and P. Faber. 2017. "Verb collocations for natural disasters: A contrastive study of Mexican and Peninsular Spanish". *Sendebar* 28: 189-208.

Buendía Castro, M., Montero Martínez, S. and P. Faber. 2014. "Verb collocations and phraseology in EcoLexicon". *Yearbook of Phraseology* 5 (1): 57-94.

Busse, D. 2012. *Frame-Semantik: Ein Kompendium*. Berlin: Walter de Gruyter.

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

Coseriu, E. 1977. *Principios de semántica estructural*. Madrid: Gredos.

Dalpanagioti, T. 2013. "Frame-semantic issues in building a bilingual lexicographic resource. A case study of Greek and English motion verbs". *Constructions and Frames* 5 (1): 1-34.

Evert, S. 2009. "Corpora and collocations". *Corpus Linguistics. An International Handbook. Volume 2*. Eds. A. Lüdeling and M. Kytö. Berlin: Walter de Gruyter. 1212-1248.

Ferraresi, A., Zanchetta, E., Baroni, M. and S. Bernardini. 2008. "Introducing and evaluating ukWaC, a very large web-derived corpus of English". *Proceedings of the 4th Web as Corpus Workshop (WAC-4)– Can we beat Google? Marrakech, 1 June 2008*. Eds. S. Evert, A. Kilgarriff and S. Sharoff. <http://wacky.sslmit.unibo.it/lib/exe/fetch.php?media=papers:wac4_2008.pdf>. (Accessed April 4 2019).

Fillmore, C. 2007. "Valency issues in FrameNet". *Valency. Theoretical, Descriptive and Cognitive Issues*. Eds. T. Herbst and K. Götz-Votteler. Berlin and New York: Mouton de Gruyter. 129-160.

Fillmore, C., Johnson, C. R. and M. R. L. Petruck. 2003. "Background to FrameNet". *International Journal of Lexicography* 16 (3): 235-250.

Fontenelle, T. 2009. "A bilingual database for Frame Semantics". *Multilingual FrameNets in Computational Lexicography. Methods and Applications*. Ed. H. C. Boas. Berlin: Mouton de Gruyter. 37-57.

Gries, S. T. 2013. "50-something years of work on collocation: What is or should be next...". *International Journal of Corpus Linguistics* 18 (1): 137-165.

Hausmann, F. J. 1979. "Un dictionnaire des collocations est-il possible?". *Travaux de linguistique et de litterature* 17 (1): 187-195.

Hausmann, F. J. 1997. "Semiotaxis und Wörterbuch". *Linguistische Theorie und lexikographische Praxis: Symposium's Vorträge, Heidelberg 1996*. Eds. K.-P. Konerding and A. Lehr. Berlin: Mouton. 171-180.

Hausmann, F. J. 1998. "O diccionario de colocacións. Criterios de oganización". *Actas do I Coloquio Galego de Fraseoloxía (1997)*. Coord. X. Ferro Ruibal. Santiago de Compostela: Centro Ramón Piñeiro para a Investigación en Humanidades. 63-81.

Hausmann, F. J. 1999. "Semiotaxis and learners' dictionaries". *The Perfect Learners' Dictionary (?)*. Ed. Th. Herbst and K. Popp. Tübingen: Max Niemeyer. 205-211.

Hausmann, F. J. 2004. "Was sind eigentlich Kollokationen?". *Wortverbindungen – mehr oder weniger fest*. Ed. K. Steyer. Berlin: de Gruyter. 309-334.

Hausmann, F. J. 2007 (2006). "Lexicographie française et phraséologie". *Collocations, phraséologie, lexicographie: Études 1977-2007 et Bibliographie*. Ed. E. Haag. Aachen: Shaker. 121-153.

Heid, U. 1994. "On ways words work together – Topics in lexical combinatorics". *EURALEX 94 Proceedings*. Eds. W. Martin, W. Meijs, M. Moerland, E. Pas, P. van Sterkenburg and P. Vossen. Amsterdam. 226-257. <http://www.euralex.org/elx_proceedings/Euralex1994/>. (Accessed 23 November 2016).

Herbst, T., Heath, D., Roe, I. F. and D. Götz. 2004. *A Valency Dictionary of English. A Corpus-based Analysis of the Complementation Patterns of English Verbs, Nouns and Adjectives*. Berlin and New York: Mouton de Gruyter.

Írsula Peña, J. 1994. *Substantiv-Verb-Kollokationen. Kontrastive Untersuchungen Deutsch-Spanisch*. Frankfurt a. M.: Peter Lang.

Johnson, M. and A. Lenci. 2011. "Verbs of visual perception in Italian FrameNet". *Constructions and Frames* 3 (1): 9-45.

Jones, S. and J. Sinclair. 1974. "English lexical collocations. A study in computational linguistics". *Cahiers de Lexicologie* 24: 15-61.

Kilgarriff, A. and D. Tugwell. 2001. "WORD SKETCH: Extraction and display of significant collocations for lexicography". *Information Technology Research Institute Technical Report Series*. <https://www.kilgarriff.co.uk/Publications/2001-KilgTugwell-ACLcollos-Sketches.pdf>. (Accessed 1 December 2016).

Koike, K. 2001. *Colocaciones léxicas en el español actual: estudio formal y léxico-semántico*. Madrid: Universidad de Alcalá / Takushoku University.

Langacker, R. 1987. *Foundations of Cognitive Grammar. Volume I: Theoretical Prerequisites*. Stanford, CA: Stanford University Press.

Langacker, R. 1991. *Foundations of Cognitive Grammar. Volume II: Descriptive Application*. Stanford, CA: Stanford University Press.

Langacker, R. 2002 (2001). *Concept, Image, and Symbol: The Cognitive Basis of Grammar*. Berlin and New York: Mouton de Gruyter.

Longman Dictionary of Contemporary English. 2014. Harlow: Longman.

L'Homme, M.-C., Laneville, M.-E. and D. Azoulay. 2014. "DiCoEnviro. Le dictionnaire fondamental de l'environment". Observatoire de Linguistique Sens-Texte. <http://olst.ling.umontreal.ca/dicoenviro/manuel-DiCoEnviro.pdf>. (Accessed 1 April 2019).

Macmillan English Dictionary for Advanced Learners. 2002. Ed. Michael Rundell. Oxford: Macmillan.

Martin, W. 2003. "Definitions and collocations in dictionaries: The GWDS compared to the Van Dale Groot Woordenboek der Nederlandse Taal". Lexicographica, Series Maior 113: 3-23.

Martin, W. 2008. "A unified approach to semantic frames and collocational patterns". *Phraseology. An interdisciplinary perspective*. Eds. S. Granger and F. Meunier. Amsterdam: John Benjamins. 51-65.

Mason, O. 2000. "Parameters of collocation: The word in the centre of gravity". *Corpora Galore. Analyses and Techniques in describing English*. Ed. J. M. Kirk. Amsterdam: Rodopi. 267-280.

Mel'čuk, I. 1998. "Collocations and lexical functions". *Phraseology. Theory, Analysis and Applications*. Ed. A. P. Cowie. Oxford: Clarendon Press. 23-53.

Mel'čuk, I. 2003. "Les collocations: définition, rôle et utilité". *Les collocations: analyse et traitement*. Eds. F. Grossman and A. Tutin. Amsterdam: De Werelt. 23-31.

Mel'čuk, I., Clas, A. and A. Polguère. 1995. *Introduction a la lexicologie explicative et combinatoire*. Louvain-la-Neuve: Duculot.

Montero Martínez, S. and M. Buendía Castro. 2017. "Clasificación semántica de colocaciones verbales para la adquisición y codificación de conocimiento experto. El caso de los riesgos naturales". *Revista Española de Lingüística aplicada* 30 (1): 240-272.

Nerima, L., Wehrli, E. and V. Seretan. 2010. "A recursive treatment of collocation". *Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC'10), Valleta, Malta*. Eds. N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner and D. Tapias. 634-638. <http://www.lrec-conf.org/proceedings/lrec2010/>. (Accessed 2 November 2016).

Oxford Dictionary of English. 2010. Ed. Angus Stevenson. Oxford: Oxford University Press.

Oxford Thesaurus of English. 2006. Ed. Maurice Waite. Oxford: Oxford University Press.

Ruppenhofer, J., Baker, C. F. and C. J. Fillmore. 2002. "Collocational information in the FrameNet database". *Proceedings of the Tenth EURALEX International Congress, EURALEX 2002, Copenhagen, Denmark, August 13-17. Vol. 1*. Eds. A. Braasch and C. Povlsen. 359-369. <http://www.euralex.org/elx_proceedings/Euralex2002/037_2002_V1_Josef%20Ruppenhofer,%20Collin%20F.%20Baker%20&%20Charles%20J.%20Fillmore_Collocational%20Information%20in.pdf>. (Accessed 28 March 2018).

Ruppenhofer, J., Ellsworth, M., Petruck, M. R. L., Johnson, C. R., Baker, C. F. and J. Scheffczyk. 2016. *FrameNet II: Extended Theory and Practice*. <https://framenet2.icsi.berkeley.edu/docs/r1.5/book.pdf>. (Accessed 28 March 2018).

Rychlý, P. 2008. "A lexicographer-friendly association score". *Proceedings of Recent Advances in Slavonic Natural Language Processing, RASLAN 2008*. Eds. P. Sojka and A. Horák. Brno: Masaryk University. 6-9. <https://nlp.fi.muni.cz/raslan/2008/papers/13.pdf>. (Accessed 5 March 2017).

Sánchez Rufat, A. 2010. "Apuntes sobre las combinaciones léxicas y el concepto de colocación". *Anuario de Estudios Filológicos* 33: 291-306.

Schumacher, H., Kubczak, J., Schmidt, R. and V. de Ruiter. 2004. *VALBU – Valenzwörterbuch deutscher Verben*. Tübingen: Gunter Narr.

Seretan, V. 2011. *Syntax-based Collocation Extraction*. Heidelberg: Springer.

Siepmann, D. 2005. "Collocation, colligation, and encoding dictionaries. Part I: Lexicological aspects". *International Journal of Lexicography* 18 (4): 409-443.

Sinclair, J. 1991. *Corpus, Concordance, Collocation*. Oxford University Press.

Stubbs, M. 2002. *Words and Phrases. Corpus Studies of Lexical Semantics*. Oxford: Blackwell.

Sullivan, K. 2013. *Frames and constructions in metaphoric language*. Amsterdam and Philadelphia: John Benjamins.

Tutin, A. 2008. "For an extended definition of lexical collocations". *Proceedings of the 13th EURALEX International Congress. Barcelona 15-19 July 2008*. Eds. E. Bernal and J. De Cesaris. Barcelona: Institut Universitari de Lingüística Aplicada. 1453-1460. <http://www.euralex.org/proceedings-toc/euralex_2008/>. (Accessed 3 November 2016).

Uhrig, P. and T. Proisl. 2012. "Less hay, more needles: Using dependency-annotated corpora to provide lexicographers with more accurate lists of collocation candidates". *Lexicographica* 28: 141-180.

Vincze, O. and M. Alonso Ramos. 2013. "Incorporating frequency information in a collocation dictionary: Establishing a methodology". *Procedia – Social and Behavioral Sciences* 95: 241-248.

Williams, G. 2003. "Les collocations et l'école contextualiste britannique". *Les collocations: analyse et traitement*. Eds. F. Grossmann and A. Tutin. Amsterdam: De Werelt. 33-44.