

Original Research Paper

Register-Specific Collocational Constructions in English and Spanish: A Usage-Based Approach

^{1,2}Gloria Corpas Pastor

¹Department of Translation Studies, University of Malaga, Spain

²RILP, University of Wolverhampton, UK

Article history

Received: 02-05-2015

Revised: 14-05-2015

Accepted: 14-05-2015

Abstract: Constructions are usage-based, conventionalised pairings of form and function within a cline of complexity and schematisation. Most research within Construction Grammar has focused on the monolingual description of schematic constructions: Mainly in English, but to a lesser extent in other languages as well. By contrast, very little constructional analyses have been carried out across languages. In this study we will focus on a type of partially substantive construction from the point of view of contrastive analysis and translation which, to the best of our knowledge, is one of the first studies of this kind. The first half of the article lays down the theoretical foundations of the study and introduces Construction Grammar as well as other formalisms used in literature in order to provide a construal account of collocations, a pervasive phenomenon in language. The experimental part describes the case study of V NP collocations with disease/enfermedad in comparable corpora in English and Spanish, both in the general domain and in the specialised medical domain. It is provided a comparative analysis of these constructions across domains and languages in terms of token-type ratio (constructional restriction-rate), lexical function, type of determiner, frequency ranking of the verbal collocate and domain specificity of collocates, among others. New measures to assess construal bondness will be put forward (lexical filledness rate and individual productivity rate) and special attention will be paid to register-dependent equivalent semantic-functional counterparts in English and Spanish and mismatches.

Keywords: Construction Grammar, Corpus Linguistics, Phraseology, Collocations, Translation

Introduction

The term *Construction Grammar* (CXG) refers to a family of closely related grammars (or constructionist approaches) which are in contrast to Chomskyan's views on language and idiomaticity. CXG aims at providing a comprehensive account of language that explains all aspects of a speaker's knowledge about their language. As Hoffmann (2013: 326) puts it, "Construction Grammar does not only focus on the idiosyncratic periphery of language, but [that] is a full-fledged grammatical theory that it is observationally, descriptively as well as explanatorily adequate".

Within constructional approaches, linguistic knowledge is acquired through repeated experience with actual instances of constructions and their generalisations. Usage frequency appears to play a crucial role in the emergence of constructions or cognitive schemata (mental representations) and their storage strength in memory (entrenchment).

Contrary to mainstream Generative Grammar, constructions are not the result of a limited set of transformations or derivations, but symbolic units which are linked to each other and constitute complex networks (Goldberg, 1995; 2006). Thus, language is conceived as

an idiomatic continuum of which constructions are the building blocks and general patterns and idioms stand on an equal footing (Kay and Fillmore, 1999: 1).

In this study we will adopt a constructionist approach to collocation in line with CXG postulates. In addition, we will take an observational stance toward a corpus-based analysis of constructions in English and Spanish, with a special emphasis on equivalent semantic-functional counterparts and potential mismatches.

This paper is organised as follows. Section 2 (On the Construal Nature of Collocations) provides a theoretical framework for collocation within competing linguistic approaches. As we will argue, collocations have a clear construal nature, as evidenced by their lexical filledness restrictions, semantic dependency and coercion along a cline of complexity and schematisation. Section 3 (Constructions in Contrast: A Case Study) furnishes a fine-grained characterisation of the V NP construction with *disease/enfermedad* in two comparable reference and specialised medical corpora in English and Spanish. We will be mainly concerned with the slot profile and its restricted lexical filledness, their productivity rates, as well as the equivalent semantic-functional counterparts and any asymmetries. The final section presents conclusions and outlines venues of further research.

On the Construal Nature of Collocations

Constructions are usage-based conventionalised pairings of form and (semantic or discourse) function that are deeply rooted in the language (entrenched). Constructions include morphemes (e.g., EN *mis-*, *-ed*; ES *des-*, *-azo*), words (e.g., EN *bank*, *to*; ES *casa*, ‘house’), filled idioms (e.g., EN *spill the beans*, ES *olerse la tostada*, ‘to smell a rat’), partially filled idioms (EN *twist sb. around/round one’s little finger*, ES *tirar de la lengua a alguien*, ‘make sb. talk’), partially lexically filled phrasal patterns (e.g., the covariational conditional construction: EN *the Xer... the Yer*; ES *cuanto más X más Y*) and fully general phrasal patterns, such as the passive construction (e.g., *She was diagnosed with breast cancer*; *Fue diagnosticada de cáncer de mama*) or the ditransitive construction (*She bought me a pen*; *Me compró un bolígrafo*).

While most cognitive grammarians would consider constructions the above examples of idiosyncratic pairings, truth is that the some distinctive features of the concept remain debatable. For example, in Cognitive Construction Grammar (CCG), frequency appears to be a determining factor: “Any linguistic pattern is recognised as a construction as long as some aspect of its form or function is not strictly predictable from its component parts or from other constructions recognised to exist. In addition, patterns are stored as

constructions even if they are fully predictable as long as they occur with sufficient frequency.” (Goldberg, 2006: 5). In the same vein, Hoffmann (2013: 315) argues that “the higher the input frequency of a particular construction, the stronger it is going to be entrenched in the neural network.”

If frequency, usage and entrenchment are valid criteria, collocations should be considered constructions in their own right.

A Cline of Complexity and Schematicity

Collocation is a pervasive, usage-based phenomenon in all languages. This fact has been extensively acknowledged in the literature. For instance, Altenberg (1998) reported that 80% of the words of the London-Lund Corpus formed part of multiword expressions. Biber and Conrad (1999) found out that between 30 and 45% of spoken English and 21% of academic prose are composed of lexical bundles. Sag *et al.* (2002: 2) claimed that specialised domain vocabulary “overwhelmingly consists of MWEs”. Jackendoff (2007) estimated that the number of multiword expressions in a language (the phrasal lexicon) equates that for single words. Finally, according to Seretan (2011), collocations represent the higher proportion of recurrent word combinations or multiword expressions (phraseology) in corpora.

Collocations appear to be a significant part of a language’s vocabulary. However, more conservative approaches, like Kay (2013), following Fillmore (1997), focus on productivity irrespectively of frequency and distinguish between fully productive processes, i.e. proper constructions, as exemplified by the *all-cleft* construction (e.g., *All I want is to hear the truth*) and semiregular processes which constitute mere patterns of coining, as the ‘A as NP’-pattern (e.g., *free as a bird*, *dark as night*, *light as a feather*). Distinctions such as true constructions as opposed to patterns of coining would definitely leave idioms and collocations outside any constructionist account of language, as they are not fully productive and, therefore, redundant.

By contrast, usage-based constructionist approaches, such as Cognitive Construction Grammar (Goldberg, 1995; 2006; 2013) and Radical Construction Grammar (Croft, 2001; 2013), consider frequency as a distinctive feature and posit, instead, a cline of constructional complexity and schematicity. Croft and Cruse (2004) distinguish atomic versus complex constructions along the complexity cline and substantive versus schematic constructions along the schematicity cline. Atomic constructions, i.e., individual morphemes, words and word classes cannot be further divided into meaningful parts, as opposed to complex constructions like *spill the*

beans (paraphraseable as ‘reveal’ + ‘information’) and *olerse la tostada* (‘suspect’ + ‘deceit’) or the Passive construction [SUBJ aux VP_{PP} (PP_{by})]. Constructions that only contain slots to be filled by various elements are said to be schematic (e.g., word class constructions and abstract constructions), whereas substantive constructions are fully lexically specified.

Constructions can be thus classified according to the different degrees of complexity and schematisation they exhibit. Such a flexible, gradient framework allows idioms and collocations to count as constructions in their own right.

From a CxG perspective, collocations can be viewed as partially specified constructions (collocational constructions) that are semantically predictable, contain slots to be filled by a restricted set of lexical items and span various phrasal patterns. Hausmann’s (1989) typology covers six main collocational types (grammar patterns): (1) noun + adjective (*célibataire endurci*, ‘confirmed bachelor’); (2) noun (subject) + verb (*la colère s’apaise*, ‘the anger wares off’); (3) verb + noun (object) (*tenir un journal*, ‘keep a diary’); (4) verb + adverb (*exiger énergiquement*, ‘insist firmly on something’); (5) adverb + adjective (*gravement malade*, ‘seriously ill’); (7) noun + (prep.) + noun (*marché du travail*, ‘labour market’). Some types would encompass further grammar patterning; e.g., pattern 1 would also cover prepositional objects (*jump to conclusions*, *poner a prueba* ‘test’); while pattern 2 includes nouns in premodification as well (*copycat criminal*, *viaje relámpago* ‘flash trip’) (Corpas Pastor, 1996; 2015).

Restricted Lexical Filledness

Unlike most idioms, collocations are not syntactically or lexically frozen (substantive constructions), nor are they completely schematic constructions. For instance, the idiomatic construction *hold one’s horses*, in the sense of slow down or curb one’s impetuosity, exhibits fixed lexical filledness (*hold one’s mules). The same lexical and morphosyntactic frozenness can be found in the Spanish construction *meter a alg. las cabras en el corral* (‘to frighten sb.’; lit. ‘get someone’s goats in the corral/get goats enter in the corral for someone’). Both idioms are non-compositional, as they have a unitary meaning, which is figurative or metaphorical.

By contrast, collocational constructions are semantically predictable but formally unpredictable (arbitrary and usage-based). Collocations pose no problem in language comprehension, provided speakers are familiar with the individual meanings of the lexical items. Let us take the examples *take/have a break* and *izar una bandera* (‘raise/hoist a flag’). Their meanings are transparent since they are the joint contribution of the individual lexical fillers in the verbal and nominal slots and yet, the actual word

combinations cannot be ascertained *a priori* as preferred ways to convey those meanings in the language. In this respect, they could be considered as encoding idioms (cf. Fillmore *et al.*, 1998).

In fact, collocations appear to be particularly problematic in language production. This is why they are so difficult to master by language learners. By way of illustration, what generalisation rule would determine that the deverbal noun *break* collocates with the lexical support verbs *take* and *have*, but not with *make* (*make a break)? In the same vein, why does the noun *bandera* (‘flag’) collocate with the verbs *izar* (specific collocation) and *subir* (generic collocation) with the meaning of ‘hoist/raise a flag’, but not with the synonymous verbs *alzar* or *levantar*? Furthermore, what rule determines the opposite cases of *make a suggestion*, but not *do/take a suggestion?; or *subir/levar ancla* (‘weigh anchor’, ‘set sail’), not *izar* (*izar un ancla), but all *izar/subir/alzar vela* (‘raise the sail’, ‘set sail’)?

Usage promotes and ensures collocational entrenchment, while exposure to those types of constructions permits speakers acquire precise knowledge about the constraints affecting the lexical filling of slots. Collocational constructions involve one or more flexible slots, which exhibit preferences for a certain restricted semantic set of lexical items. It should be borne in mind, though, that the degree of restriction is variable and goes along a cline of least restricted, such as *run* + nouns denoting ‘business’ (*hotel, bar, restaurant, shop*, etc.) to more restricted, as *shrug one’s shoulders* or *commit suicide*.

To make the point clear, let us provide some more examples. The adjective slot of the nominal collocational construction in the sense of ‘typical or persistent’ shows preferences for certain lexical items according to the noun slot. With *bachelor*, the adjective slot tends to be filled by *confirmed, hardened, hopeless, inveterate, incorrigible* and *stubborn*; whereas the adjective slot in the construction with *drunker* would usually be filled by *inveterate* or *incorrigible* and to a lesser extent, *confirmed*, but not **stubborn* or **hardened* (Corpas Pastor, 2015).

Similar lexical constraints can be observed in the case of the nominal collocational construction with the adjective *prolijo* (‘comprehensive’) in Spanish. Certain lexical fillers are attracted in the noun slot to denote (a) documents or information (*información, informe, dato, documento...*); (b) enumeration, series, catalogue (*enumeración, relación, antología, recuento...*); (c) study and research processes (*investigación, análisis, razonamiento, reflexión...*); (d) tasks and cronograms (*plan, programa, tarea, labor, trabajo...*) and (e) personal experience and achievements (*recorrido, carrera, biografía, currículum vitae...*), among others. In

the case of constructions with the synonym *exhaustivo*, similar lexico-semantic sets are selected (*estudio, trabajo, reflexión, argumentación*, etc.), but only *exhaustive* attracts lexical fillers such as *revisión, seguimiento, comprobación* or *planteamiento* in the noun slot, as stated in REDES (Bosque, 2004).

Halliday (1966) already acknowledged this phenomenon when he extended the notion of collocation (usual or habitual co-occurrence) to cover collocational restrictions and lexical sets as well: While *powerful* and *strong* can collocate with *argument*, only *strong* can collocate with *tea* (*powerful tea). This implies a gradation in collocability among the set of words which are found to occur in the environment of a particular item. Three decades later, Sinclair (1996) put forward a comprehensive, corpus-driven model of analysis for identifying and describing lexical items as *extended units of meaning*. The model is composed of five categories of co-selection: (i) Core, that is the word(s) which are invariable and always present (for instance, *naked eye*); (ii) collocation (co-occurrence of words with the core, e.g., *see, visible, invisible, apparent, evident, obvious, undetectable* at L3 (Ln means n positions to the left of the node in a KWIC line; likewise, Rn means n positions to the right) and to a lesser extent at L4); (iii) colligation (co-occurrence of grammar choices with the core, e.g., *the* at L1 position and *with, to, by, from, as, upon, tan* at L2); (iv) semantic preference, that is, the restriction of regular co-occurrence to words which share a common semantic feature (e.g., about 'vision' or 'visibility'); and (v) semantic prosody, in other words, the overall functional meaning of the lexical item, e.g., 'difficulty', also confirmed by the collocational range of *see* (*small, faint, weak, difficult*) and *visible* (*barely, rarely, just* and modal verbs *can, could*).

The model of extended unit of meaning (bottom-up) seems to be compatible with the notion of partially specified constructions (top-down).

Semantic Dependency and Coercion

In this subsection we will make reference to relevant semantic and formal features of collocations, such as dependency, predictability, typicality, coercion, bondness and bipartite structure. Collocational constructions are semantically predictable. However, their symbolic nature prevents them from being entirely compositional. Or quite the opposite: Precisely because of their construal nature they affect the meaning contribution of their specific slots fillers. As Goldberg (1995: 4) puts it, "systematic differences in meaning between the same verb in different constructions are attributed directly to the particular constructions". Beside the usage-based

preferences which affect the lexical filling of slots, there is, in fact, certain semantic dependency within the constituents in given collocation (see Hausmann, 1979; 1985; 1989; 2007; Benson *et al.*, 1986; Benson, 1989; Bartsch, 2004). Collocations usually exhibit a bipartite structure, conventionally restricted, in which both collocates possess a different semantic status: One of them is the semantically autonomous word (the base) and the other is the semantically dependent component (the collocate). Complex collocations like *pretty horrible weather* or *recibir duras críticas* ('be criticized harshly') are actually embedded bipartite entities: [[pretty, horrible] weather], [recibir[duras, críticas]] (See Heid, 1994, on the recursive nature of collocations).

While bases keep their primary meanings, collocates usually select specialised senses (delexical, secondary or figurative), as in *pay attention/prestar atención; gain strength/cobrar fuerza; cherish a hope/acariciar una idea*; and *raise suspicion/levantar una sospecha*, among others.

In language production, bases select collocates in a unidirectional fashion. In other words, the selection of a collocate is contingent upon the prior selection of the base. This is especially noticeable across languages and in translation. By default, nouns tend to be the bases (e.g., *complaint* selects the verb *file*), except for patterns 5 and 6 where, in the absence of nouns, verbs and adjectives are the bases (e.g., *deny* selects *strongly*). This shows a certain prominence of NPs in general and of nouns as syntactic heads of NPs, an insight which has been retained by the CXG models, Hausmann (1979; 1985; 1989; 2007) has contributed some of the most influential ideas in the advancement of the semantic approach to collocations. Other relevant contributions are Benson *et al.* (1986; Benson, 1989; Bartsch, 2004), especially chapters 2-4.

Base-collocate unidirectional dependency has been somehow formalised in other linguistic traditions. Within the Meaning-Text Theory (MTT) (Melčuk, 1973; 1996; Melčuk and Pertsov, 1987; Wanner, 2007, among others), the notion of *lexical function* (LF) is used in the mathematical sense $f(X) = Y$ in order to describe and represent semantic relationships among words. For instance, if 'INTENSIFICATION' were to be considered a function (f_i), then the semantic relation among the constituents of the collocation *keenly aware* could be represented as $f_i(\text{ARGUMENT}) = \text{VALUE}$: $f_i(\text{aware}) = \text{keenly}$. If (f_i) was applied to other arguments in type 1, the resulting values would possibly vary as well: $f_i(\text{apologetic}) = \text{deeply}$; $f_i(\text{unusual}) = \text{highly}$; $f_i(\text{despicable}) = \text{utterly}$; $f_i(\text{sound}) = \text{asleep}$. (Actually, there is an LF intensifier: Magn (Lat. *magnus*), e.g. 'intense(ly)', 'very', 'to a (very) high degree'.)

But the unidirectional semantic dependency found in collocations could also be interpreted in terms of typicality, usage and frequency (e.g., *hoist a flag/izar una bandera; fell a tree/italar un árbol*); or even as a case of bonded coercion (*explode a myth/acabar con un mito; lift a sanction/levantar una sanción*).

In Cognitive Linguistics, terms like *coercion*, *accommodation*, *enriched composition* or *implicit conversion* refer to the interpretative adjustment which happens in case of a conflict “between the semantic properties of a selector (be it a construction, a word class, a temporal or aspectual marker) and the inherent semantic properties of a selected element, the latter no being expected in that particular context” (Lawers and Willems, 2011: 49). This would explain the existence of collocational constructions where collocates are forced to be reinterpreted in a grammaticalised or figurative fashion by their selectors (bases). Let us take the case of the collocation *explode a myth*. The verb *explode*, which usually selects concrete nouns (physical objects) with the potential to burst or shatter violently and noisily (e.g., *bomb, Molotov cocktail*), co-occurs with the abstract noun *myth*. This mismatch within the causative construction triggers the metaphorical interpretation of the verb as ‘show something to be false or no longer true’ and widens the semantic set of lexical fillers to nouns denoting ‘misconception’ (*belief, idea, notion, theory*). Similarly, *cobrar* (‘get/collect money’) prototypically selects concrete nouns related to payments, wages and debts (*dinero, sueldo, salario, factura, deuda, comisión*, etc.). However, when it takes as complement abstract nouns denoting positive qualities metaphorically related to brightness (*resplandor, brillo, lustre, esplendor*), the verb meaning undergoes grammaticalisation.

This type of unidirectional coercion is akin to the systemic concept of *meaning by collocation* (Firth, 1957, 1968). It would explain, then, (a) the delexicalisation or metaphorisation of collocates when there is semantic mismatch with their selectors, as well as the (b) bondness coercion of collocational constructions in general and in relation to the co-selection of fillers.

Within CXG models, collostructional analysis explores the bondness (association strength) within constructions and their slot fillers. It comprises “a family of quantitative corpus-linguistic methods for studying the relationship between words and the grammatical structures they occur in” (Stefanowitsch, 2013). Simple collexeme analysis detects significantly attracted collexeme of a given constructions, while distinctive collexeme analysis compares the association strength of all collocates of two near-

synonymous constructions (e.g., the ditransitive and the prepositional dative constructions). Finally, the covarying collexeme analysis is a structure-sensitive collocate analysis which takes into account the syntactic relation of the words within a given construction.

Taylor (2002) argues that coercion brings bondness and that the strength of the bond depends on the degree of alteration of the element coerced by its neighbours in the construction. As a result, components are caused to develop fuzzy boundaries which affect their individual identities and give rise to new, extra meanings.

In any case, it should be borne in mind that (a) the bondness observed for collocational constructions is brought about by frequency, usage and semantic dependency, whether based on unidirectional selection and/or coercion; and (b) bondness brings entrenchment and vice versa.

Constructions in Contrast: A Case Study

Research on contrastive construction grammar has attracted insufficient attention. Despite its title, the volume edited by Fried and Östman (2004) does not include contrastive analyses of languages pairs from a CxG viewpoint, but constructionist analyses for English, Czech, Japanese and French in a monolingual fashion.

An exception which confirms the rule is the volume edited by Boas (2010). It contains contrastive analysis in CxG for English and Swedish/Spanish/Russian/Finnish/Thai/Japanese, among others in a number of corpus-based chapters which illustrate the so-called Contrastive Construction Grammar (CCxG).

In the case of Spanish-English, most studies so far have focused on syntax. In the edited volume by Boas just mentioned, González-García’s (2010) paper deals with the accusative cum infinitive after verbs of cognition and communication in small clauses in English and Spanish. In a recent paper (González-García, 2014), also in a volume co-edited with Boas (2010; Boas and González-García, 2014), the author studies the semantic-pragmatic restrictions in secondary predication in English and Spanish involving right- or left-dislocation. In the same volume, Pedersen (2014) examines verb-framed constructions of telic motion with manner verbs in Spanish and compares them with English, a satellite-framed language.

In this study we will focus on V NP constructions of the type of partially substantive constructions studied by Wulff (1998; 2013) for English. We will adopt a usage-based CCxG perspective on collocational constructions in English and Spanish which, to the best of our knowledge, is one of the first studies of this kind. Other contrastive contributions in

the field of phraseology can be found in this special issue of the *Journal of Social Sciences*.

Choice of Corpora and Domains

In order to illustrate our approach, we will concentrate on V NP collocational constructions with slots fillers of the medical domain. This is a particularly interesting field, with terms coexisting in restricted/specialised registers (medical discourse) and in common uses of the language (general discourse). For example, constructs such as *remove a cyst* and *hacer/realizar una radiografía*, ‘take an X-ray’ tend to appear frequently in both general and specialised medical corpora. Unlike domain-specific technical terms (used only by specialists), banalised terms are domain core-terms which are found in both domain-specific genres and in general, mainstream communication.

In this study, it is understood that *register* "is the set of meanings, the configuration of semantic patterns, that are typically drawn upon under the specific conditions, along with the words and structures that are used in the realization of these meanings" (Halliday, 1978: 23). Therefore, the term *register* refers to language variation according to use and encompasses the context of situation in which the communicative event takes place, as well as the typical linguistic features associated with it and its varying values: field (purpose activity, domain), mode (spoken/written, genre) and tenor (social role, participant's interaction).

Our assumption is that V NP constructions with terms from the health domain as slot fillers will show some substantial differences depending on whether they are used in specialised medical discourse or in a context of general communication. If this is the case, V NP constructions could be considered register-specific. This is in line with Kerz and Wiechmann's (2015) postulates. After identifying patterns of adverbial clause constructions in academic and journalist registers, the authors argue that constructions have register-specific entrenchment values.

The register-specificity of V NP constructions could also accommodate a traditional claim within the systemic tradition on collocation. In fact, the term *collocation* was initially introduced by Firth (1957; 1968) to mean (a) a mode of semantic analysis or ‘meaning by collocation’ (cf. examples above) and (b) a stylistic means to characterise restricted languages and levels of formality, as illustrated by *take away parental rights* (general language, neutral) versus *terminate parental rights* (specialised legal English, formal); or *poner una multa* (general language, neutral) versus *imponer una multa* (specialised legal/administrative Spanish, formal) and *casaca una multa* (general language, informal), which all mean ‘to give someone

a fine’ in a cline of formality and specialisation (See also relevant research on lexical bundles as a means to characterise registers and language varieties (cf. Biber, 1995; Biber and Conrad, 1999; Cortes, 2004; Connor and Upton, 2004; Hyland, 2008; Chen and Baker, 2010, etc.). Lexical bundles are continuous sequences of two or more words (e.g., on the other hand, can I have a, in the case of the) which are retrieved from corpora according to a specified frequency threshold, regardless of their meanings and their structural status).

Space restrictions prevent us from providing a detailed account of all V NP constructions of the medical domain in Spanish and English. Therefore a small set of collocational constructs have been selected for this case study. They are licensed by a V NP construction whose noun slot fillers convey the concept of ‘ailment’ in both languages: *disease* \approx *enfermedad*. This choice is justified because both lexical items have a high frequency of occurrence in both general and specialised corpora. In addition, they provide a convenient *tertium comparationis* for the analysis, given that they prima-facie translation equivalents.

The following corpora were used for the analysis presented in this study:

BNC-BYU-the web-version of the 100 million word *British National Corpus* (1970s-1993) (Davies 2004-) (<http://corpus.byu.edu/bnc/>).

CORPES XXI-the *Reference Corpus of 21st Century Spanish* (RAE), a pan-Spanish general corpus of over 170 million words (1975-2014) (<http://web.frl.es/CORPES/view/inicioExterno.view>). It contains the Reference Corpus of Contemporary Spanish (CREA, <http://corpus.rae.es/creanet.html>) and updates. The current beta version (0.7) allows more flexible access to the data (lemmatisation, morphological disambiguation, collocations, statistics), although the system is still rather unstable and slow in terms of processing. Another drawback is that the corpus is under construction, which means that results may vary significantly according to the access date. The CORPES is expected to reach over 500 million words in 2018) from which the European or Castilian subcorpus has been selected (30% of all documents).

TELLME-the *TELL-ME Medical Corpus* (*English, Spanish and German*), a trilingual comparable corpus of over 20 million words of medical discourse, from which the Spanish and English subcorpora have been selected (TELLME_ES and TELLME_EN). A detailed description of the corpus can be found in Gutiérrez Florido *et al.* (2013).

Occurrences have been retrieved using AntConc (v. 3.4.3.) and manually checked, whereas the general, reference corpora have been analysed by means of their own in-built corpus management and retrieval systems.

Our approach involves intra- and interlinguistic analysis. The intralinguistic analysis is actually a two-step process involving an initial study of collocational constructions licensed for the selected slot fillers in mainstream discourse, which is followed by a comparison of the results obtained in the specialised medical corpus. The interlinguistic analysis compares the former collocational constructions across languages.

Verbal Slot Fillers for V NP Constructions with Disease in English

Within the BNC-BYU corpus, the V NP collocational constructions with *disease* (V NP_{disease}) exhibit a wide range of verbal slot fillers, e.g. *contract* DET *disease*, *have* DET *disease*, *catch* DET *disease*, *cause* DET *disease*, *develop* DET *disease*, etc. Table 1 lists all verbal slot fillers for V NP_{disease} in the BNC-BYU corpus, with their frequencies (in their word forms or types) in parenthesis. The Constructional Restriction Rate (CRR) of the verbal slot fillers in a particular construction can be calculated by dividing the total number of the number of tokens (word forms occurrences) by types (word forms). In the case of V NP collocational constructions with *disease* in general English, the restriction rate is 1.88 (CRR = 314/167).

In theory DET stands for any kind of determiner, including a zero determiner. But in the case of *disease*, it only takes the form of definite determiner (*the*), which shows certain anaphoric preferences (ex. 1) (Examples 1-2 have been extracted from the BNC-BYU corpus). There are no cases with zero or other types of determiners (*a*, *any*, *some*, *one*, etc.), unless the noun takes a premodifier (ex. 2):

(1) "One of these AIDS victims was a haemophiliac who probably contracted the disease through infected blood products."

(2) "Each day they grew in number and I began to get really worried, wondering if, while in Greece, I had caught a contagious disease."

The verbal slot fillers tend to convey the main collocational meanings associated to type 3 collocations, according to Benson *et al.* (1986): (I) Creation and/or activation, including spreading of the disease (e.g., *contract*, *catch*, *spread*, *cause*, *develop*, *get*, *take*, *carry*, *trigger*, *transmit*, *pass*, *re-awaken*, *reproduce*, *incubate*, *reveal*, etc.); and (II) eradication and/or nullification, including treatment and control (e.g., *control*, *prevent*, *cure*, *combat*, *stop*, *treat*, *overcome*, *eradicate*, *halt*, *fight*, etc.). Notice the coercion exerted by *disease* on the construal interpretation of many verbal slot fillers, such as *fight*, *combat*, *battle*, *beat*, *encourage*, *feed*, *hide*, *rescue*, etc.

Other frequent verbal collocates have to do with the actual state of being ill or having been diagnosed with a disease, facing or evaluating the condition (*have*, *diagnose*, *suffer*, *detect*, *assess*, *embrace*), among others. (Other frequent constructions, such as VPass constructions (*be diagnosed with a disease*) or V PP constructions (*suffer from a disease*, *afflicted with a disease*) are beyond the scope of this study). Finally, some slot fillers are closely linked with specific contexts, such as AIDs (*deserve* DET *disease*) or rehabilitation programs for Parkinson (*delay* DET *disease*).

A cursory look at V NP_{disease} in the specialised medical corpus shows a rather different picture. On the one hand, DET usually takes the form of the indefinite determiner (*a*), usually followed by a *that/which*-clause or *such as* (ex. 3) or else zero determiner in the presence of a premodifier (ex. 4). The indefinitive determiner reveals a preferential cataphoric use of the NP *disease* in the specialised medical discourse, as a structuring device for topicalisation.

(3) Using PGD to ensure a baby does not carry an altered gene which would guarantee a baby would inherit a disease such as cystic fibrosis, is well-established.

(4) So a lot of women have had Graves' disease 10 years prior to a pregnancy.

Occasionally DET is realised by the definite determiner *the* or the demonstrative *this*, as in ex. 5-6, which in fact have been extracted from the public domain part of TELLME_EN.

(5) People can transmit the disease as long as the bacteria remain in their system.

(6) There was no treatment, no vaccine to prevent contracting this disease; the doctor prescribed injections of vitamin B12.

On the other hand, the range of verbal slot fillers that enter into V NP constructions with *disease* is much more restricted (33 types): *catch* DET *disease*, *cause* DET *disease*, *contract* DET *disease*, *detect* DET *disease*, *diagnose* DET *disease*, *get* DET *disease*, *have* DET *disease*, *inherit* DET *disease* and *transmit* DET *disease* (see all word types in Table 2), being *have* ('actual estate of being ill'), *develop* and *cause* ('creation/activation') the most frequent collocates. Some slot fillers, like *accumulate*, *aggravate*, *counter* or *survey* are specific of the specialised medical domain and cannot be found in the general corpus. In addition, V PP constructions (e.g., *die from a disease*, *linked with a disease*) tend to appear in the TELLME corpus as well.

Table 1. Verbal slot fillers (BNC-BYU)

contract (22), have (19), cause (12), catch (13), develop (10), control (8), diagnose (6), prevent (6), spread (6), carry (5), cure (7), pass (4), fight (4), treat (4), stop (3), transmit (3), combat (2), assess (2), confront (2), detect (2), eradicate (2), get (2), halt (2), identify (2), investigate (2), suffer (2), overcome (2), precipitate (2), reproduce (2), slow (2), trigger (2), understand (2), accept (1), acquire (1), battle (1), beat (1), check (1), counteract (1), confirm (1), contain (1), define (1), deserve (1), determine (1), delay (1), embrace (1), encourage (1), evaluate (1), experience (1), fear (1), feed (1), hide (1), ignore (1), incubate (1), inherit (1), initiate (1), introduce (1), link (1), manifest (1), match (1), mistake (1), modulate (1), realise (1), re-awaken (1), recall (1), reduce (1), reflect (1), reproduce (1), rescue (1), reveal (1), share (1), study (1), spell (1), spot (1), surround (1), tackle (1), watch (1), worsen (1)

Table 2. Verbal slot fillers (TELLME EN)

have (159), develop (53), cause (34), treat (28), get (20), prevent (20), diagnose (8), contract (7), control (6), detect (6), reduce (6), transmit (6), spread (5), cure (5), worsen (5), aggravate (4), carry (4), identify (4), catch (2), confirm (2), fight (2), incubate (2), inherit (2), manifest (2), study (2), contain (1), counter (1), determine (1), eradicate (1), initiate (1), pass (1), survey (1), tackle (1)

The constructional restriction rate of the verbal slot fillers in the specialised corpus is much higher than in the case of general English: 12.24 (CRR = 404/33). This comes to show that the range of eligible verbal slot fillers in the specialised medical corpus is more restricted and shows less variation to express similar construal meanings (activation, eradication, experiencing the condition, etc.). Therefore, the bond between the noun and the verbal slot fillers licensed by the V NP_{disease} in specialised medical discourse is stronger. On the other hand, the coercion exerted by the noun is milder in this case, since there is a lower number of verbs whose primary meaning needs to be reinterpreted, as in the case of *fight* or *incubate*.

Interestingly enough, while other general verbal slot fillers (e.g., *trigger*, *reflect*, *overcome*, etc.) do not enter into V NP constructions with the lexical item *disease*, they do so with its cohyponyms: *Trigger sinusitis*, *reflect the disease mechanism*, *overcome spinal cord paralysis*, etc. Some of the general verbal slots also enter with *disease* and its cohyponyms in other types of constructions, such as nominalisations (e.g., *assessment of disease*) or attributive constructions with participial adjectives (e.g., *a slowly progressing disease*, *defined genetic diseases*).

Specialised verbal slot fillers like *aggravate*, specific to the medical corpus, do not enter in constructions with *disease* in the BNC-BYU, but with partial synonyms (*condition*, *injury*, *damage*, *illness*). Similarly, synonymous verbal slot fillers behave differently: *counter* appears to be restricted to specialised medical English, whereas *counteract* tends to appear in the general corpus instead.

However, the main differences do not affect so much the actual list of lexical fillers, but rather (a) their position in the frequency rank, (b) their percentage of lexical filledness and (c) their individual productivity rate (Table 1 and 2). The Lexical Filledness Rate (LFR) refers to the probability of co-occurrence of a particular slot filler within a given construction. It is calculated by means of cross-multiplication: individual number of tokens multiplied by 100 and divided by total number of tokens for the

slot filler in a given construction. The Individual Productivity Rate (IPR) is calculated by dividing the number of tokens of a given slot filler by the number of total tokens for a particular slot within a given construction. LFR and IPR are inversely proportional: the higher the LFR, the lower the IPR and vice versa.

For instance, *contract* is ranked 1 in BNC-BYU, with an IPR of 14.27 (IPR = 314/22) and a LFR of 7 (LFR = [22×100]/314). In TELLME_EN it appears in the 8th position of the rank, with an IPR of 57.71 (IPR = 404/7) and a LFR of 17.32. On the other hand, while *tackle* is ranked 33 in the medical corpus and 12 in the general corpus, the corresponding rates do not vary that much: IPR 314 (IPR) and 0.31 (LFR) in BNC-BYU and IPR = 404 (IPR) and 0.24 (LFR) in the medical corpus.

Verbal Slot Fillers for V NP Constructions with Enfermedad in Spanish

The CORPES XXI is composed of a European or Castilian subcorpus (30%) and an American subcorpus (70%). The Castilian subcorpus contains approximately 51 million words (half the size of the BNC-BYU corpus). However, the range of verbal slot fillers for V NP_{enfermedad} is proportionally smaller: 30 Spanish types versus 167 in English (almost 5 times less).

For the purpose of this study, only transitive constructions have been taken into account within a collocational window span of 3. Other constructions (e.g., ditransitive, causative, passive, etc.) exhibiting the same collocational patterns and intransitive uses (e.g., *riesgo de + V NP_{enfermedad}*), verbs which can have transitive and intransitive uses are *adquirir(se)*, *agravar(se)*, *contagiar(se)*, *contraer(se)*, *desarrollar(se)*, *extender(se)*, *propagar(se)*, or *transmitir(se)*, among others. Other frequent constructions which convey 'tendency to', 'predisposition to' are *susceptible de/propenso a/capaz de/predisposición a + V NP_{enfermedad}*.

The most frequent Spanish general V NP collocational constructs are *prevenir DET enfermedad*, *curar DET enfermedad*, *sufrir DET enfermedad*, *padecer DET enfermedad*, *diagnosticar DET enfermedad*, *transmitir DET enfermedad*, *detectar DET enfermedad*

and *evitar* DET *enfermedad*. Their raw frequencies are all above 30 (Table 3). The constructional restriction rate of the verbal slot fillers for V NP_*enfermedad* in CORPES XXI is 26.28 (CRR = 841/32).

Most verbal slot fillers in V NP_*enfermedad* convey the main two construal meanings associated with this construction: activation (*contraer*, *transmitir*, *contagiar*, etc.) and eradication (*curar*, *superar*, *erradicar*, etc.), as well as other typical meanings related to the detection and treatment of the illness and individual ways to face the situation (*diagnosticar*, *afrontar*, *manifestar*, etc.). Some lexical fillers exhibit specialised, metaphorical senses as a result of the coercion exerted by the noun on their construal interpretation. For example, *combatir* (combat) and *ahuyentar* (frighten away) in their primary senses usually take animated, living beings as complements; however, their construal meanings have been coerced by the noun (conceived metaphorically as an enemy) into metaphorical and figurative interpretations ('fight against this enemy and keep it at bay').

Other verbs listed in the Spanish combinatorial dictionary REDES (Bosque, 2004) as typical verbal collocates for *enfermedad* do not actually appear as slot fillers of the V NP_*enfermedad* construction. This applies to *incubar*, *arrastrar*, *combatir*, *bregar*, *lidiar*, *pillar*, *mitigar*, *contrarrestar*, *coger*, etc. However, they tend to collocate with sets of hyponyms (ex. 7-8), similarly to the phenomenon also observed in the BNC-BYU corpus:

(7) *combatir* + *la pérdida de densidad ósea/el estrés/ el VIH/la depresión*

(8) *pillar* + *una hepatitis/una neumonía/una pulmonía/un resfriado/un catarro/lumbago*

Finally, some of the verbal slot fillers appear to exert influence also on the DET and/or show preferences as to the grammatical number of the noun within the construction. For instance, *padecer* and, to lesser extent, *contraer* tend to select zero determiner and/or plural noun ("padecer/enfermedades"); *aliviar* usually selects the 3rd person possessive adjective *su*; *agrarar* shows a tendency to select the noun in plural, while others, like *curar*, appear to be quite unrestrictive as to the number of the noun and the DET selected: zero; indefinite, definitive, demonstrative and possessive determinants in singular (*la*, *esta*, *aquella*, *cierta*, *una*, *su*, etc.) and plural (*muchas*, *diversas*, *las*, *unas*, *tus*, *todas*, *tantas*, etc.).

In the case of specialised medical Spanish, the list of slot fillers decreases to 25 types and 547 tokens in the TELLME_EN corpus, which results in a slightly reduced CRR (21.88). The most frequent constructions in the corpus are *padecer* DET *enfermedad*, *sufrir* DET *enfermedad*, *tener* DET *enfermedad*, *desarrollar* DET *enfermedad*, *prevenir* DET *enfermedad* and *presentar* DET *enfermedad* (over 20 co-occurrences, see Table 4).

A substantial number of occurrences in the corpus realise DET as a zero determiner before the noun in plural, or else as definite or demonstrative determiner of the noun in singular or plural.

Again, the main types of construal meanings can be observed: activation (e.g., *adquirir*), eradication (e.g., *atajar*), detection and treatment (e.g., *diagnosticar*), etc. Coercion is much less widespread in specialised medical Spanish, although it is present in *atajar* ('take a shortcut', 'reach a person or animal by taking a shortcut'), whose construal meaning is closer to the 'eradicate' group.

There are verbal slot fillers for V NP_*enfermedad* which only appear in the general corpus, not in the specialised medical corpus (*ahuyentar*, *combatir*, *contagiar*, *declarar*, *extender*, *pegar*, *propagar*, *vencer*). Some of them (*contagiar*, *manifestar*, *propagar*) do appear in TELLME_ES inside other types of constructions (usually nominalisations) with the same collocational pattern (ex. 9). In addition, some verbal slot fillers are mainly restricted to specialised medical domains, such as *gestionar* and *esconder* (ex. 10).

(9) Apareció hace ya 20 años en el mercado, es más seguro que el masculino para proteger del contagio de enfermedades de transmisión sexual y su fiabilidad contra el embarazo iguala la del condón tradicional.

(10) De acuerdo a la visión de los especialistas consultados, los gases son un mero trastorno y no tienen por qué esconder una enfermedad.

A cursory look at the actual lists of verbal slot fillers for V NP_*enfermedad* in both corpora show differences as regards the selection of slot fillers, their number of types and tokens, their relative positions in both frequency ranks and their percentage of lexical filledness. For instance, *prevenir* is ranked 1 in CORPES XXI, with an IPR of 8.08 (IPR = 841/104) and a LFR of 12.36 (LFR = [104×100]/841); whereas it appears no. 5 in the specialised medical corpus, with an IPR of 23.78 (IPR = 547/23) and a LFR of 4.22. In the same vein, *padecer*, which is the top verb in the specialised medical corpus (IPR = 2.49; LFR = 40.03), is ranked 5 in the general corpus (IPR = 13.34; LFR = 7.49). A similar mismatch can be seen as regards *curar*, *diagnosticar* or *transmitir*, among many others.

It is worth noting that V NP medical constructions of the type just analysed do not only differ interlinguistically (an evidence of their register-specificity), but also from a contrastive, intralinguistic perspective. Let us illustrate our point in relation to the construal meaning 'eradication' and its verbal realisations in the general and specialised corpora.

Table 3. Verbal slot fillers (CORPES XXI)

prevenir (104), curar (100), contraer (84), sufrir (65), padecer (63), diagnosticar (55), transmitir (50), detectar (40), evitar (33), causar (29), tener (25), superar (24), desarrollar (22), presentar (22), producir (21), combatir (16), agravar (15), contagiar (12), erradicar (12), aliviar (9), estudiar (8), propagar (7), originar (6), vencer (4), atajar (3), declarar (3), pegar (3), adquirir (2), afrontar (1), ahuyentar (1), extender (1), manifestar (1)

Table 4. Verbal slot fillers (TELLME_ES)

padecer (219), sufrir (76), tener (58), desarrollar (48), prevenir (23), presentar (22), contraer (15), detectar (15), superar (15), evitar (8), curar (7), causar (7), adquirir (5), agravar (5), producir (5), estudiar (4), diagnosticar (3), transmitir (2), erradicar (2), gestionar (2), originar (2), afrontar (1), aliviar (1), atajar (1), esconder (1)

Table 5. Verbal slot fillers for ‘eradication’

V NP _{disease}		V NP _{enfermedad}	
BNC-BYU	TELLME EN	CORPES XXI	TELLME ES
beat	control	curar	aliviar
battle	cure	ahuyentar	atajar
check	counter	aliviar	erradicar
combat	eradicate	atajar	superar
confront	reduce	combatir	
contain		erradicar	
control		evitar	
counteract		superar	
cure		vencer	
eradicate			
fight			
halt			
overcome			
modulate			
reduce			
slow			
stop			

The trends observed in Table 5 are quite representative of the entire list of verbal slot fillers in both languages. The greatest difference can be seen in the list of plausible verbs, which is most restricted in the case of specialised medical Spanish, followed by medical specialised English. The V NP_{enfermedad} exhibits a noticeably different list of verbal slot fillers in both general and specialised medical Spanish. However, both varieties do not vary as much as their English counterparts for the V NP_{disease}. These results are compatible with the Constructional Restriction Rate (CRR) found in general English versus specialised medical English: 1.88/12.24; and in general Spanish versus specialised medical Spanish: 26.28/21.88. The CRR marks the stylistic distance within a language and between languages: medical Spanish is closer to general Spanish, than general English to medical English. The difference in the first case is 10.36, but only 4.4 for Spanish.

Regarding the individual slot fillers, some prima-facie equivalents (*cure* ≈ *curar*) may appear in the general corpus, but not in the specialised one (e.g., *curar* does not appear in the TELLME_ES corpus). In any

case, even when they were present in the two corpora for both languages, they would still vary as to their ranking, their percentage of Lexical Filledness (LFR) and their Individual Productivity Rate (IPR) and possibly their pragmatic features (implicatures, negative/positive evaluation, etc.).

One final remarkable difference affects the presence of V NP and NP collocational constructions with the slot filler *disease* and *enfermedad*. In the case of English, the BNY-BYU corpus appears to contain a greater proportion of V NP constructions, whereas the TELLME_EN corpus appears to favour nominalisation, with a clear predominance of NP constructions (*artery/arterial disease, a sexually transmitted disease, non-alcoholic fatty liver disease*) over V NP constructions. This suggests a tendency towards nominalisation in specialised medical discourse, which is not so evident in the case of Spanish. While this observation needs to be tested using further quantitative research, it could nevertheless be incorporated in a vector of features for establishing the degree of specialisation and formality within medical genres and across languages.

Conclusion

Collocations constitute a type of partially specified constructions that are semantically predictable but formally unpredictable (i.e., collocational constructions). They involve one or more flexible slots which exhibit preferences for a certain restricted semantic set of lexical items, along a cline of complexity and schematicity.

Usage promotes collocational entrenchment and the acquisition of the actual constraints affecting slot lexical filledness. Slots fillers in collocational constructions play a very important semantic and bonding role. They interact with the construction as a whole, but probably also among themselves. The constructs licensed by a particular collocational construction will depend on various types of semantic dependency and bonded coercion, as well as on generalisations based on repeated exposure to those constructs.

Collocational constructions are register-specific, as suggested by the case study presented in this paper on V NP constructions in the medical domain. The V NP_{disease} and V NP_{enfermedad} constructions exhibit relevant differences per type of corpus and language. Our findings will be especially relevant as most CxG research has focused on the monolingual description of constructions (mainly English), while very little work has been conducted across languages.

So far, collocation analysis has explored the association strength within constructions and their slot fillers (cf. also syntax-based collocation extraction). However, no attention has been paid to the degree of internal restriction observed within the construction as a whole and the verbal slot fillers licensed by the construction across languages and registers.

A number of parameters have been suggested in this study to complement present-day collocation analyses along the lines mentioned. The Constructional Restriction Rate (CRR) measures the intrinsic bondness of a given construction across different registers and genres. For instance, the V NP_{disease} construction in the specialised medical corpus has a CRR of 12.24, as opposed to a CRR of 1.88 in the general corpus. This means that the range of eligible slot fillers in the English medical corpus is more restricted and exhibits less variation for the expression of the same construal meanings. By contrast, the Spanish equivalent construction V NP_{enfermedad} shows less distance among registers (only a 4.4 difference).

As to the individual relationships among lexical items that realise a particular slot within the same construction, the following parameters have been put forward for each lexical item within a given construction: (a) their position in the frequency ranks of different corpora, (b) their Lexical Filledness Rate (LFR) or co-occurrence

probability and (c) their Individual Productivity Rate (IPR), which equates low values with high productivity and vice versa. Those parameters can be applied to perform intralinguistic analyses (for example, of a slot filler or sets of synonymous slot fillers across registers or genres) and interlinguistic analyses (for example, of equivalent slots fillers across languages and/or registers). The results could shed light on typological and variational aspects of languages.

The results presented in this study have a very narrow scope. It would be interesting to extend the analysis to more V NP constructs within the 'ailment' domain, such as synonymous constructions (e.g., V NP_{disease} vs. V NP_{illness}), hyperonymous constructions V NP_{disease} vs. V NP_{fibromyalgia}), both from a monolingual and a bilingual perspective. Also the degree of complexity could be another interesting line of research (e.g., constructions with simple and complex slot fillers: *suffer, battle*, etc. Vs. *suffer from, nip in the bud or ascribe to*), as well as comparing different types of schematic constructions which attract similar lexical items as slot fillers (e.g., passive, ditransitive construction, nominalisation, etc). Again these types of analyses could be performed in a monolingual fashion, within registers and/or across languages.

Acknowledgement

I would like to thank two anonymous reviewers for their valuable comments and suggestions to improve the paper.

Funding Information

The research presented in this study has been partially carried out in the framework of research projects Expert (317471-FP7-PEOPLE-2012-ITN), Tradicor (PIE 13-054) and Inteliterm (FFI2012-38881).

Ethics

This article is original and contains unpublished material. The author confirms that the manuscript has been read and approved and no ethical issues are involved.

References

- Altenberg, B., 1988. On the Phraseology of Spoken English: The Evidence of Recurrent Word-Combinations. In: *Phraseology: Theory, Analysis and Applications*, Anthony, P.C. (Ed.), Oxford University Press, Oxford, pp: 101-122.
- Bartsch, S., 2004. Structural and functional properties of collocations in English: A corpus study of lexical and pragmatic constraints on lexical co-occurrence. Narr, Tübingen.

- Benson, M., 1989. The structure of the collocational dictionary. *Int. J. Lexicography*, 2: 1-3.
DOI: 10.1093/ijl/2.1.1
- Benson, M., E. Benson and R. Ilson, 1986. *The BBI Combinatory Dictionary of English: A Guide to Word Combinations*. 1st Edn., John Benjamins, Amsterdam, ISBN-10: 9027220379, pp: 286.
- Biber, D., 1995. *Dimensions of Register Variation: A Cross-Linguistic Comparison*. 1st Edn., Cambridge University Press, Cambridge, ISBN-10: 0521473314, pp: 428.
- Biber, D. and S. Conrad, 1999. Lexical Bundles in Speech and Writing: An Initial Taxonomy. In: *Corpus Linguistics by the Lune: A Festschrift for Geoffrey Leech*, Wilson, A., P. Rayson and T. McEnery (Eds.), Peter Lang, Frankfurt, pp: 71-92.
- Boas, H.C., 2010. *Contrastive Studies in Construction Grammar*. 1st Edn., John Benjamins Publishing, Amsterdam, ISBN-10: 9027204322, pp: 244.
- Boas, H.C. and F. González-García, 2014. *Romance Perspectives on Construction Grammar*. 1st Edn., John Benjamins Publishing Company, ISBN-10: 9027204373, pp: 326.
- Bosque, I., 2004. *Redes: Diccionario Combinatorio del Español Contemporáneo*. 2nd Edn., Ediciones SM, Madrid, ISBN-10: 8467502762, pp: 1839.
- Chen, Y.H. and P. Baker, 2010. Lexical bundles in L1 and L2 academic writing. *Lang. Learn. Technol.*, 14: 30-49.
- Connor, U. and T.A. Upton, 2004. *Discourse in the Professions: Perspectives from Corpus Linguistics*. 1st Edn., John Benjamins Publishing, Amsterdam, ISBN-10: 9027222878, pp: 333.
- Corpas Pastor, G., 1996. *Manual de Fraseología Española*. 1st Edn., Gredos, Madrid, ISBN-10: 8424918290, pp: 337.
- Corpas Pastor, G., 2015. Collocations in e-Bilingual Dictionaries: from Underlying Theoretical Assumptions to Practical Lexicography and Translation Issues. In: *Collocations and other lexical Combinations in Spanish. Theoretical and Applied Approaches*, Torner, S. and E. Berbal (Eds.), Ohio State University Press, Chicago, IL.
- Cortes, V., 2004. Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English Specific Purposes*, 23: 397-423.
DOI: 10.1016/j.esp.2003.12.001
- Croft, W.A., 2001. *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. 1st Edn., Oxford University Press, Oxford, ISBN-10: 9780198299547, pp: 448.
- Croft, W., 2013. *Radical Construction Grammar*. In: *The Oxford Handbook of Construction Grammar*, Hoffmann, T. and G. Trousdale (Eds.), Oxford University Press, Oxford, pp: 211-232.
- Croft, W.A. and D.A. Cruse, 2004. *Cognitive Linguistics*. 1st Edn., Cambridge University Press, Cambridge, ISBN-10: 1139445227.
- Davies, M. 2004. *BYU-BNC. Based on the British National Corpus from Oxford University Press*.
- Fillmore, C.J., 1997. *Construction grammar lecture notes*.
- Fillmore, C.J., P. Kay and M.C. O'Connor, 1988. Regularity and idiomaticity in grammatical constructions: The case of *let alone*. *Language*, 64: 501-38.
- Firth, J.R., 1957. *Papers in Linguistics, 1934-1951*. 1st Edn., Oxford University Press, London, pp: 233.
- Firth, J.R., 1968. *Linguistic Analysis as a Study of Meaning*. In: *Selected Papers of J. R. Firth 1952-59*, Palmer, F.R. (Ed.), Longmans, Londres, pp: 12-26.
- Fried, M. and J.O. Östman, 2004. *Construction Grammar in a Cross-Language Perspective*. 1st Edn., John Benjamins Publishing, Amsterdam, ISBN-10: 9027218226, pp: 208.
- Goldberg, A.E., 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. 1st Edn., University of Chicago Press, Chicago, ISBN-13: 9780226300863, pp: 271.
- Goldberg, A.E., 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford University Press, New York, ISBN-10: 0199268517, pp: 280.
- Goldberg, A.E., 2013. *Constructionist Approaches*. In: *The Oxford Handbook of Construction Grammar*, Hoffmann, T. and G. Trousdale (Eds.), Oxford University Press, Oxford, pp: 15-31.
- González-García's, F., 2010. *Contrasting Constructions in English and Spanish: The Influence of Semantic, Pragmatic and Discourse Factors*. In: *Contrastive Studies in Construction Grammar*, Boas, H.C. (Ed.), John Benjamins, Amsterdam/Philadelphia, pp: 43-86.
- González-García, F., 2014. *Bringing Together Fragments and Constructions: Evidence from Complementation in English and Spanish*. In: *Romance Perspectives on Construction Grammar. (Constructional Approaches to Language 15)*, Boas, H.C. and F. González-García, (Eds.). John Benjamins, Amsterdam/Philadelphia, pp: 181-226.
- Gutiérrez Florido, R., G. Copas Pastor and M. Seghiri, 2013. Using semi-automatic compiled corpora for medical terminology and vocabulary building in the healthcare domain. *Proceedings of the 10th International Conference on Terminology and Artificial Intelligence, (TAI' 13)*, Université Paris 13, Villetaneuse, France, pp: 14-18.
- Halliday, M.A.K., 1966. *Lexis as a Linguistic Level*. In: *Memory of John Firth*, Bazell, C.E., J.C. Catford, M.A.K. Halliday and R.H. Robins (Eds.), Longman, London, pp: 148-162.

- Halliday, M.A.K., 1978. *Language as a Social Semiotic: The Social Interpretation of Language and Meaning*. 1st Edn., University Park Press, Baltimore, ISBN-10: 0839111835, pp: 256.
- Hausmann, F.J., 1979. Un dictionnaire des collocations est-il possible? *Travaux de Linguistique et de Littérature*, 17: 187-195.
- Hausmann, F.J., 1985. Kollokationen in deutschen Wörterbuch: Ein Beitrag zur Theorie des lexikographischen Beispiels. In: *Lexikographie und Grammatik: Akten des Essener Kolloquiums zur Grammatik in Wörterbuch*, (Lexicographica Series Maior 3), Bergenholtz, H. and J. Mugdam (Eds.), Max Niemeyer, Tübingen, pp: 118-129.
- Hausmann, F.J., 1989. Le Dictionnaire de Collocations. In: *Wörterbücher. Dictionaries. Dictionnaires. Eininternationales Handbuch zur Lexikographie. An International Encyclopedia of Lexicography. Encyclopédie Internationale de Lexicographie*, Hausmann, F.J., O. Reichmann, H.E. Wiegand and L. Zgusta (Eds.), Walter de Gruyter, Berlin/New York, pp: 1000-1019.
- Hausmann, F.J., 2007. Die Kollokation im Rahmen der Phraseologie-systematische und historische Darstellung. *Zeitschrift für Anglistik und Amerikanistik*, 55: 217-234.
- Heid, U., 1994. On ways words work together-research topics in lexical combinatorics. *Proceedings of the 6th Euralex International Conference Congress on Lexicography*, (CCL' 94), Euralex, Amsterdam, pp: 226-257.
- Hoffmann, T., 2013. Abstract Phrasal and Clausal Constructions. In: *The Oxford Handbook of Construction Grammar*, Hoffmann, T. and G. Trousdale (Eds.), Oxford University Press, Oxford, pp: 307-328.
- Hyland, K., 2008. As can be seen: Lexical bundles and disciplinary variation. *English Specific Purposes*, 27: 4-21.
- Jackendoff, R., 2007. *Language, Consciousness, Culture: Essays on Mental Structure*. 1st Edn., The MIT Press, Cambridge, Mass., ISBN-10: 026210119X, pp: 403.
- Kay, P., 2013. The Limits of (Construction Grammar). In: *The Oxford Handbook of Construction Grammar*, Hoffmann, T. and G. Trousdale (Eds.), Oxford University Press, Oxford, pp: 32-48.
- Kay, P. and C.J. Fillmore, 1999. Grammatical constructions and linguistic generalizations: the *What's X doing Y?* Construction. *Language*, 75: 1-33. DOI: 10.1353/lan.1999.0033
- Kerz, E. and D. Wiechmann, 2015. Register-contingent entrenchment of constructional patterns: Causal and concessive adverbial clauses in academic and newspaper writing. *J. English Linguist.*, 43: 61-85. DOI: 10.1177/0075424214564364
- Lawers, P. and D. Willems, 2011. Coercion: Definition and challenges, current approaches and new trends. *Linguistics*, 49: 1219-1235. DOI: 10.1515/ling.2011.034
- Melčuk, I., 1973. Towards a Linguistic 'Meaning Text' Model. In: *Trends in Soviet Theoretical Linguistics*, Kiefer, F. (Ed.), Reidel, Dordrecht, pp: 33-57.
- Melčuk, I., 1996. A Tool for the Description of Lexical Relations in the Lexicon. In: *Lexical Functions in Lexicography and Natural Language Processing*, Wanner, L. (Ed.), John Benjamins, Amsterdam/Philadelphia.
- Melčuk, I. and N.V. Pertsov, 1987. *Surface Syntax of English: A Formal Model within the Meaning-Text Framework*. 1st Edn., John Benjamins, Amsterdam, ISBN-10: 9027215154, pp: 526.
- Pedersen, J., 2014. Variable Type Framing in Spanish Constructions of Directed Motion. In: *Romance Perspectives on Construction Grammar. (Constructional Approaches to Language 15)*, Boas, H.C. and F. González-García, (Eds.), John Benjamins, Amsterdam/Philadelphia, pp: 269-304.
- Sag, I.A., T. Baldwin, F. Bond, A. Copestake and D. Flickinger, 2002. Multiword expressions: A pain in the neck for NLP. *Proceedings of the 3rd International Conference on Computational Linguistics and Intelligent Text Processing*, Feb. 17-23, Springer, Mexico, pp: 1-15.
- Seretan, V., 2011. *Syntax-Based Collocation Extraction*. 1st Edn., Springer, Dordrecht, ISBN-10: 9400701349, pp: 232.
- Sinclair, J.M., 1996. The search for units of meaning. *TEXTUS*, 9: 75-106.
- Stefanowitsch, A., 2013. Collostructional Analysis. In: *The Oxford Handbook of Construction Grammar*, Hoffmann, T. and G. Trousdale (Eds.), Oxford University Press, Oxford, pp: 290-306.
- Taylor, J.R., 2002. *Cognitive Grammar*. 1st Edn., Oxford University Press, Oxford, ISBN-10: 0198700334, pp: 634.
- Wanner, L., 2007. Selected Lexical and Grammatical Issues in the Meaning-Text Theory: In honour of Igor Melčuk. 1st Edn., John Benjamins, ISBN-10: 9027230943, pp: 377.
- Wulff, S., 1998. *Rethinking Idiomaticity: A Usage-based Approach*. 1st Edn., Continuum, New York.
- Wulff, S., 2013. Words and Idioms. In: *The Oxford Handbook of Construction Grammar*, Hoffmann, T. and G. Trousdale (Eds.), Oxford University Press, Oxford, pp: 274-289.