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

Gloria Corpas Pastor

Department of Translation Studies, University of Malaga, Spain
RIILP, University of Wolverhampton, UK

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 explanatory 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 cuarto más X más Y) and fully general phrasal patterns, such as the passive construction (e.g., She was diagnosed with breast cancer; Fue diagnostica da 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 distinct 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 posits, 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 (paraphrasable as ‘reveal’ + ‘information’) and
olerse la tostada (‘suspect’ + ‘deceit’) or the Passive
construction [SUBJ aux VPpp (PPby)]. 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’apease, ‘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 relâmpago ‘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
(meaning, which is figurative or metaphorical.

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, curriculum 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, planteamiento in the noun slot, as stated in REDES (Buscós, 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 collocatability 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).

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 Pernot, 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), then the semantic relation among the constituents of the collocation keenly aware could be represented as f_{\text{ARGUMENT}} = \text{VALUE} ; f_{\text{aware}} = \text{keenly}. If (f_{\text{fi}}) was applied to other arguments in type 1, the resulting values would possibly vary as well: f_{\text{apologetic}} = \text{deeply} ; f_{\text{unusual}} = \text{highly} ; f_{\text{despicable}} = \text{utterly} ; f_{\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 typicity, 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, collocational 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, Gonzálvez-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 (Gonzálvez-García, 2014), also in a volume co-edited with Boas (2010; Boas and Gonzálvez-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.

[This text continues with further discussions and analyses related to the study of collocations in contrastive construction grammar.]
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 slot 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 *cascara 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 = 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:

- CORPES XXI-the Reference Corpus of 21st Century Spanish (RAE), a pan-Spanish general corpus of over 170 million words (1975-2014) (http://web.frIes/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).
- TELL-ME-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 subcorpus have been selected (TELL-ME ES and TELL-ME 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-BYC 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 indefinite 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.
English, whereas of co-occurrence of a particular slot filler within a Lexical Filledness Rate (LFR) refers to the probability individual productivity rate (Table 1 and 2). The percentage of lexical filledness and (c) their position in the frequency rank, (b) their synonymous verbal slot fillers behave differently: disease to the medical corpus, do not enter in constructions with their cohyponyms: genetic diseases as in the case of verbs whose primary meaning needs to be reinterpreted, milder in this case, since there is a lower number of constructions, such as nominalisations (e.g., assessment) or attributive constructions with participial adjectives (e.g., a slowly progressing disease, defined genetic diseases).

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 simtisitis, 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

---

**Table 1. Verbal slot fillers (BNC-BYU)**

<table>
<thead>
<tr>
<th>Slot Filler</th>
<th>BNC-BYU IPR</th>
<th>BNC-BYU LFR</th>
<th>Specialised IPR</th>
<th>Specialised LFR</th>
</tr>
</thead>
<tbody>
<tr>
<td>contract</td>
<td>314</td>
<td>0.31</td>
<td>404</td>
<td>0.24</td>
</tr>
<tr>
<td>have</td>
<td>159</td>
<td>0.31</td>
<td>314</td>
<td>0.24</td>
</tr>
<tr>
<td>develop</td>
<td>10</td>
<td>0.31</td>
<td>22</td>
<td>0.24</td>
</tr>
<tr>
<td>disease</td>
<td>404</td>
<td>0.24</td>
<td>57.71</td>
<td>0.31</td>
</tr>
<tr>
<td>transmit</td>
<td>6</td>
<td>0.31</td>
<td>404</td>
<td>0.24</td>
</tr>
<tr>
<td>ces</td>
<td>57.71</td>
<td>0.31</td>
<td>0.24</td>
<td>0.31</td>
</tr>
</tbody>
</table>

---

**Table 2. Verbal slot fillers (TELLME EN)**

<table>
<thead>
<tr>
<th>Slot Filler</th>
<th>TELLME EN IPR</th>
<th>TELLME EN LFR</th>
<th>Specialised IPR</th>
<th>Specialised LFR</th>
</tr>
</thead>
<tbody>
<tr>
<td>contract</td>
<td>404</td>
<td>0.24</td>
<td>314</td>
<td>0.24</td>
</tr>
<tr>
<td>have</td>
<td>314</td>
<td>0.24</td>
<td>404</td>
<td>0.24</td>
</tr>
<tr>
<td>develop</td>
<td>22</td>
<td>0.24</td>
<td>57.71</td>
<td>0.31</td>
</tr>
<tr>
<td>disease</td>
<td>57.71</td>
<td>0.31</td>
<td>0.24</td>
<td>0.31</td>
</tr>
<tr>
<td>transmit</td>
<td>404</td>
<td>0.24</td>
<td>57.71</td>
<td>0.31</td>
</tr>
<tr>
<td>ces</td>
<td>314</td>
<td>0.24</td>
<td>404</td>
<td>0.24</td>
</tr>
</tbody>
</table>

---

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_[enfermedad] 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.
and evitár 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, combatír (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) combatír + la pérdida de densidad ósea/el estrés/ el VIH/la depresión

(8) pillar + una hepatitis/una neumonia/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; agravar 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_Eng corpus, which results in a slightly reduced CRR (21.88). The most frequent constructions in the corpus are padecír DET enfermedad, sufrír 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., adquirír), 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.
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_\text{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_\text{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 (\textit{cure} \approx \textit{curar}) may appear in the general corpus, but not in the specialised one (e.g., \textit{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 \textit{disease} and \textit{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 analyses could be performed in a monolingual fashion, (a) 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 equivalents 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 constructions within the ‘ailment’ domain, such as synonymous constructions (e.g., V NP_suffer vs. V NP_disease), 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 or 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


Davies, M. 2004. BYU-BNC. Based on the British National Corpus from Oxford University Press.


