

# Error Analysis of Support Verb Constructions in Written Spanish Learner Corpora

Marcos García Salido \*

LyS research team

Department of Galician-Portuguese, French and Linguistics

Facultade de Filoloxía, Universidade da Coruña

Campus da Zapateira, 15071, A Coruña

*Abstract:* This article studies the use of Support Verb Constructions (SVCs) in the written production of learners of Spanish. SVCs are lexical combinations made up of a verb and a noun whose content is similar to verbal predicates, but is distributed between a verb and a noun, the noun being the carrier of the core lexical meaning of the predicate. In spite of the fact that there is considerable agreement on the importance of these constructions in the learning process, their use in the production of learners of Spanish has so far attracted little attention. This study examines the difficulties posed to learners by this construction by means of a qualitative analysis of the errors registered in three samples consisting of essays by learners with three different mother tongues (English, Swedish and Japanese). It focuses particularly on three types of error: two of them (support verb choice and determiner choice) seem to be especially problematic due to the unpredictable choice of the units involved. The third type (using an SVC instead of a more idiomatic one-word verb) is regularly found in only one set of the samples (the Japanese speakers'), which suggests the influence of a particular mother tongue in its production.

*Keywords:* support verb constructions, Spanish learners, collocations, learner corpora, error analysis

---

\* This study was possible thanks to a post-doctoral fellowship granted by the Galician Government (POS-A/2013/191).

NOTICE: this is the authors' version of a piece of research that was accepted for publication in *The Modern Language Journal*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in García Salido, Marcos. 2016. Error Analysis of Support Verb Constructions in Written Spanish Learner Corpora. *Modern Language Journal*, 100(1), 362–376. DOI: 10.1111/modl.12320

## 1. INTRODUCTION

There seems to be a broad consensus among language learning researchers on the importance of collocations as a decisive factor determining the fluency and the native-like appearance of learners' linguistic production (Granger, 1998; Higuera, 2006; Howarth, 1998). This article will deal with part of the Spanish collocational repertoire: the so-called *support verb constructions* (henceforth SVCs). SVCs are collocations of the verb+noun type, the noun of which conveys a predicative content and the verb, having a weak lexical meaning, adds the grammatical features characteristic of its part-of-speech: tense, mood and person (e.g. *dar un paseo*, 'to take a walk') (M. Gross, 1981; G. Gross, 1989; Alonso Ramos, 2004a; Mel'čuk, 2004).

SVCs represent an important subset amongst the collocations of Spanish and are one of the most frequently used collocation types of this language: in the case of learners of this language, Wanner, Verlinde & Alonso Ramos (2013) have found that SVCs were the collocations most frequently used in the learner corpus they studied, accounting for 35.7% of correct collocations and 24% of incorrect ones<sup>1</sup>. However, in spite of their importance in language learning, few studies have addressed the issue of how learners of Spanish use SVCs (Martín Bosque, 2006; Buckingham, 2008; Mitatou, 2011; Molina-Plaza & de Gregorio-Godeo, 2010; Wanner, Verlinde, & Alonso Ramos, 2013; García Salido, 2014). Furthermore, only a few of them have analysed learner data (Martín Bosque, 2006; Wanner, Verlinde & Alonso Ramos, 2013; García Salido, 2014),

perhaps due to the lack of availability of resources such as learner corpora until relatively recently.

From the work that has already been done in this field two main ideas with regard to learners' use of SVCs emerge: (i) SVCs pose Spanish learners difficulties that are apparently absent in the case of one-word units (cf. Buckingham, 2008 or Molina-Plaza & de Gregorio-Godeo, 2010); and (ii) these difficulties are likely to make learners avoid SVCs in favour of their one-word counterparts, so that an underuse of these constructions is to be expected in their production.

This second idea is challenged in García Salido (2014). This article compares the frequency and lexical variety of SVCs in two samples extracted from the CEDEL2 corpus (Lozano & Mendikoetxea, 2013), one made up of learners' essays and the other consisting of native speakers' essays. From this comparison, it can be concluded that the frequency of use of SVCs is very similar in both samples, and even slightly higher in the learner sample. The lexical variety of the SVCs' repertoire of the latter is, however, lower than that observed in the native group. Thus, it seems that learners make the most of a more restricted repertoire of SVCs than that of native speakers.

As in the case of the idea that SVCs are underused by learners of Spanish, the assumption of their difficulty is mostly based on data that have not been retrieved from learner production (specialized corpora, in the case of Buckingham, 2008; reference corpora in Molina-Plaza and de Gregorio-Godeo, 2010; and dictionaries and text books in the case of Mitatou, 2011). The aim of this article is, then, to explore the difficulties posed by SVCs, if any, to learners of Spanish, by means of a study of real learner data.

The approach adopted in undertaking this task will be a qualitative one. This analysis will be carried out using data from the corpus CEDEL2, whose collocations have been annotated with a system based on Mel'čuk's Lexical Functions (Mel'čuk,

1996) and whose collocational errors have been classified according to a typology proposed by Alonso Ramos et al. (2010a, b). Besides the SVC errors annotated in the CEDEL2 corpus, data from the CORANE corpus (Cestero Mancera & Penadés Martínez, 2009) will be included. In contrast to CEDEL2, which is made up of essays from persons with an English-speaking background, CORANE includes writings of learners with different mother tongues (L1s), which can be a valuable source of data in order to determine the influence of the L1 in learners' production.

The present article is organised as follows: after this introduction, an explanation of the concept of SVC adopted here is provided. Next, I will describe the corpora that have been used and the methodology applied to identify SVCs and to establish their incorrectness. The following section lists and describes recurring errors in learners' SVCs. A discussion section follows that aims at establishing the difficulties behind the errors listed in the previous section. The article ends with the exposition of some concluding remarks.

## 2. THE CONCEPT OF SUPPORT VERB CONSTRUCTION

It has been noted by several scholars that some verbs convey a weak lexical meaning, serving mainly to lend their grammatical features (tense, mood, person) to a predicate, whose core lexical meaning is conveyed by a noun dependent on those verbs. Such verbs have been labelled as *light verbs* (Jespersen, 1946: 117), *delexical verbs* (Sinclair & Renouf, 1985) or *support verbs*, among other denominations. The latter label is particularly associated with two theoretical models that have devoted considerable attention to this phenomenon during the last decades: Lexicon-Grammar (cf., among others, G. Gross, 1989; M. Gross, 1981 for French and Blanco Escoda, 2000 for Spanish) and Meaning-Text Theory (MTT) (see, for instance, Mel'čuk, 2004 or,

focusing on Spanish, Alonso Ramos, 2004a). This article will follow the characterisation of SVCs offered by Alonso Ramos (2004a) within the MTT.

The characteristic feature of a support verb is that of being semantically empty. However, in certain contexts, verbs with a lexical content less schematic or general than that of *hacer* ('to do, to make') or *tener* ('to have') can be assimilated to this category, as their lexical content does not add new lexical features to those already provided by a noun governed by them (e.g., *cometer un crimen*, 'to commit a crime' or *decir mentiras*, 'to tell lies'). Thus, Alonso Ramos (2004a: 82ff) distinguishes between two types of lexically empty verbs: (i) those that are paradigmatically empty, that is, that have a meaning that is more general or abstract than that of other verbs, and (ii) those that are syntagmatically emptied, that is, verbs with a more complex lexical content than those in (i), but that in the context of the supported noun do not make any substantial contribution to the lexical content of the predicate. For example, *decir* ('to say, to tell') can be considered a support verb in the expression *decir mentiras* ('to tell lies') because its lexical meaning is already contained in the lexical unit (LU) *mentira* ('lie') (cf. Alonso Ramos, 2004a: 87). The support verb, therefore, is not selected by virtue of its lexical content, but as a support for the noun that carries the predicate meaning (Alonso Ramos, *ibid.*; Mel'čuk, 2004).

With respect to the noun of an SVC, there is a wide consensus in considering it the lexical core of the predicate, but the set of criteria defining a predicate noun vary depending on the framework considered: thus, according to some scholars, only "abstract" or "non-referential" nouns qualify to be considered as predicates (cf. G. Gross, 1989: 22-23; De Miguel Aparicio, 2008: 568), predicate nouns cannot have resultative interpretations (De Miguel Aparicio, 2008), etc. Following Alonso Ramos once more, this article will consider that the conclusive trait in defining the nouns of

SVCs is their character as semantic predicates. Within the MTT framework, a semantic predicate is a meaning that (i) has argument slots and (ii) denotes a fact (facts can be paraphrased as ‘something that can happen’, in contrast to entities, that are ‘something that can be’) (Polguère, 2012). Nouns fulfilling the two afore-mentioned criteria are predicates from a semantic perspective. Nouns that meet criterion (i), but not (ii), are considered “quasi-predicates” (e.g., nouns denoting a relation, such as *father*; nouns denoting instruments, such as *knife*, etc.). Both semantic predicates and semantic quasi-predicates may be syntactically governed by a support verb, giving rise to an SVC.

As pointed out in the introduction, in the present piece of research and according to the MTT framework, SVCs will be considered a subclass of collocations. It has already been said that the support verb is not selected because of its lexical meaning, which is very weak. On the contrary, its selection depends on the identity of the predicate noun. That is precisely the relation that holds between the members of a collocation: one of them, the *base*, is selected freely to convey a certain meaning. The choice of the other —the *collocate*—, however, is contingent on the lexical identity of the predicate noun (see, e.g., Mel’čuk, 2012a: 39). In the case of SVCs, if a speaker has chosen a noun such as *conversación* (‘conversation’), for instance, he will not be able to express the performance of the activity denoted by the noun combining it with the verb *hacer* (*\*hacer una conversación*, ‘to do a conversation’), which in fact means ‘to do, to make’, but with *tener* (*tener una conversación*, ‘to have a conversation’).

### 3. METHODOLOGICAL ISSUES

Two corpora have been used for the present study: the CEDEL2 (Lozano & Mendikoetxea, 2013) and the CORANE (Cestero Mancera & Penadés Martínez, 2009). The CEDEL2 contains written texts produced by Spanish learners with English as their L1 and a comparable collection of texts written by native Speakers of Spanish. It

comprises samples of all proficiency levels, but only a subpart of it, containing 103 learner texts of intermediate and advanced levels and 100 native speaker texts, has been annotated with collocations and collocational errors. The learner texts consist of ca. 52,000 tokens.

The inclusion of another corpus in the analysis is a consequence of the fact that the CEDEL2 contains only samples of learners with the same L1. Whereas the second corpus, the CORANE, lacks a section of comparable texts produced by native speakers, it does include samples of learners with different L1s, which can yield interesting results for a qualitative comparison of the errors made in collocation production. Two groups of learners from intermediate to advanced proficiency levels (B1 and C1 of the CEFR<sup>2</sup>) have been selected from the CORANE corpus: their L1s are Swedish and Japanese. The learners with these two L1s are the best represented in the levels of the CORANE chosen and, therefore, may provide an amount of data that is sufficiently representative for the purposes of the present study. The Swedish sample contains ca. 47,000 tokens and the Japanese one ca. 53,000 tokens. The number of informants in the three samples, however, varies to a greater extent, as can be seen in table 1.

CORPUS	No. of learners	Tokens
CEDEL2	103	ca. 52,000
CORANE (L1 Japanese)	38	ca. 53,000
CORANE (L1 Swedish)	26	ca. 47,000

Table 1: Samples composition

The annotation of the collocations of CEDEL2 (including the SVCs) is the result of a research project that has already concluded<sup>3</sup>. All the collocations were annotated by means of Lexical Functions (Mel'čuk, 1996), a tool that provides a semantic and syntactic description of collocations:

- a) From a semantic perspective, lexical functions describe the meaning of the collocates in the context of a given base or *keyword*. For instance, the lexical function **Magn** describes collocates that act as intensifiers whose meaning can be paraphrased by ‘intensely, very’ (see Mel’čuk, 1996). Therefore, when this function takes a noun such as *fumador* (‘smoker’) as its argument, it yields as its value the collocate *gran* (‘big, great’) for Spanish and *heavy* for English (**Magn**(*smoker*)=*heavy*).
- b) From the syntactic point of view, information regarding the syntactic structure of collocations can be derived from the lexical function assigned to a given instance. Thus, **Magn** is usually associated with adjectives or adverbs with a modificative relation with the keyword; **Oper** describes verbs that take as its main object<sup>4</sup> a predicate noun (e.g. **Oper**<sub>1</sub>(*huelga* ‘strike’)=*hacer* [‘to go/be on’, lit. ‘to make’]).

In addition, erroneous collocations have been classified following a typology of errors also presented in Alonso Ramos et al. (2010a, b). This typology includes the following information: (i) the location of the error, that is, whether the error affects the base, the collocate or the whole collocation; (ii) the description of the error, i.e., the description of lexical, grammatical or register problems affecting the collocation, etc.; and finally (iii), an explanation of the source of the error, namely whether the error is due to the influence of another language or intra-linguistic factors. The descriptive dimension is articulated in a fine-grained classification that will be dealt with in some detail in section 4 below. It must also be noted that an explanatory analysis of the collocational errors registered in CORANE has not been tackled and, only in very specific and remarkable cases has an attempt been made to establish the origin of collocational mistakes (see the discussion below).



The collocations (both correct and incorrect) of a subpart of the CEDEL2 corpus have been manually annotated by at least two different annotators by means of the Lexical Functions repertoire and following the aforementioned error typology in the case of incorrect collocations. Only those instances where the annotators agreed upon their collocational nature have been maintained in the annotation. Likewise, only those collocations upon whose incorrectness the annotators agreed have been considered collocational mistakes. The annotation of lexical functions has made it possible to retrieve only those collocations corresponding to SVCs (Oper).

For the samples extracted from the CORANE, only erroneous instances of SVCs have been manually retrieved and classified according to the error typology referred to above, except for its explanatory dimension, as already noted. After the initial classification, a second native speaker has supervised the initial annotation. Only those collocations considered erroneous by both annotators have been taken into account for the present study.

#### 4. CLASSIFICATION OF SVC ERRORS

In this section the different types of error made by learners when producing Spanish SVCs are listed. The error typology corresponds to that presented in Alonso Ramos et al. (2010a, b), already mentioned above. After the description of the errors encountered in the three samples referred to earlier, Table 2 below provides information about their frequency and distribution. For each group, two figures are given: the first stands for the tokens of each error type (frequency), whilst the second indicates the number of different informants that committed this type of error (range).

##### 4.1. *Lexical errors*

*4.1.1. Substitution errors.* The most common error affecting the LUs of a collocation is their substitution, i.e. the replacement of one or both of them by another

existing Spanish LU. As Nesselhauf (2004: 116) points out, most studies on verb+noun collocations emphasise the difficulty of choosing the right verb, probably because this is the unit whose selection is restricted by the base —and, according to this study’s data, the verb is the most frequently substituted LU in SVCs (e.g., *llenar un puesto* instead of *ocupar un puesto*, ‘to fill a position’). But as Nesselhauf also demonstrates, the noun may also have been wrongly selected (in which case we have a *base substitution*, e.g., *hacer escenarios* instead of *hacer escenas*, ‘to make scenes’), and even both noun and verb may be inappropriate in certain contexts (e.g., *gritar abusos* instead of *lanzar insultos*, ‘to shout abuse’).

4.1.2. *Word creation*. The previous category includes those cases where an LU is substituted by another existing LU. There are, however, some instances of SVC where an LU is a non-existent word in Spanish, although this kind of error seems to be rare. In all instances, this error affects the base (e.g. *tiene limitades* instead of *tiene límites*, ‘it has limits’ in the CEDEL2 corpus, *sentir la calencia* instead of *sentir el calor*, ‘to feel the heat’, in the Japanese speakers’ sample or *hacer una compenciación* instead of *dar una compensación*, ‘to compensate’, in the Swedish-speakers’ sample).

4.1.3. *Analysis errors* . Another difficulty that occurs repeatedly but only in the Japanese-speakers’ subcorpus is the so-called *analysis error*. This label refers to the production of an expression with the characteristics of an SVC that is not very idiomatic and whose content tends to be expressed by means of a one-word LU in the target language (e.g. *hacer preparación* instead of *prepararse*, ‘to prepare oneself’; *hacer la actividad profesional* instead of *trabajar*, ‘to work’, etc.).

4.1.4. *Existing collocation with another meaning*. Another error that occurs repeatedly in the three sub-corpora examined and involves the LUs of the collocation consists in the use of an existing collocation of Spanish with a meaning different from

the one it usually conveys (e.g., *tener la autoridad*, ‘to have the authority’, instead of *tener la posibilidad*, ‘to have the possibility’ or *hacer un cariño*, ‘to make an affectionate gesture’, instead of *dar cariño*, ‘to give affection’).

4.2. *Grammar errors*. The errors listed so far can be described as lexical errors, as they affect one or both of the LUs of a collocation. However, a number of features that can be classified as grammatical, such as gender or number, but also determination or prepositional government, play a role in the production of collocations as well. Among these, the choice of determiners (e.g., *tener el humor* instead of *tener humor*, ‘have a sense of humour’; *tener bebé* instead of *tener un bebé* ‘have a baby’; etc.) and prepositions governed by the collocate or the base (e.g. *estar en vacación* instead of *estar de vacaciones*, ‘be on holiday’; *tener la posibilidad a* instead of *tener la posibilidad de*, ‘be able to’) seems to be the most problematic, as they are the most repeated grammatical mistakes in the three sub-corpora examined.

The remaining grammatical types of mistake identified in the sub-corpora studied display a lower frequency than the previous two and in some cases are completely absent from one of the samples. They either consist of the wrong assignment of number or gender to the predicate noun (e.g., *estar en vacación [de vacaciones]*, ‘to be on a holiday’; *tener muchos tiempos libres [mucho tiempo libre]*, ‘to have a lot of spare time’ *tener muchas [muchos] problemas*, ‘to have a lot of problems’; *tener el [una] conversación*, ‘to have a conversation’) or have to do with inadequate choices of the clitics accompanying the verb of the SVC, for instance, the presence of a reflexive marker where it is not expected, as in the following examples (the latter type occurs repeatedly only in the Japanese speakers’ sample).

1. ...es odio que uno **se** siente por algo (lit. ‘is hate that one REFL-feels for something’) (CEDEL2 AD)
2. ... lo que sí **me** recuerdo son memorias muy difusas (lit. ‘what I do REFL-remember are very diffuse memories’) (CORANE C1 080)

Error type		Sample	Frequency/Range
L E X I C A L	Collocate substitution	CEDEL2	33/24
		CORANE (Jap.)	18/9
		CORANE (Swed.)	16/10
	Base substitution	CEDEL2	9/8
		CORANE (Jap.)	-
		CORANE (Swed.)	8/6
	Collocate+Base substitution	CEDEL2	9/8
		CORANE (Jap.)	1/1
		CORANE (Swed.)	5/5
	Word creation	CEDEL2	1/1
		CORANE (Jap.)	2/2
		CORANE (Swed.)	2/2
	Analysis	CEDEL2	1/1
		CORANE (Jap.)	12/9
		CORANE (Swed.)	1/1
	Correct collocation with another meaning	CEDEL2	4/4
		CORANE (Jap.)	2/2
		CORANE (Swed.)	2/2

G R A M M A T I C A L	Erroneous preposition	CEDEL2	28/23
		CORANE (Jap.)	16/9
		CORANE (Swed.)	9/7
	Erroneous determiner	CEDEL2	13/10
		CORANE (Jap.)	23/13
		CORANE (Swed.)	11/9
	Number	CEDEL2	7/6
		CORANE (Jap.)	5/5
		CORANE (Swed.)	-
	Gender	CEDEL2	7/6
		CORANE (Jap.)	1/1
		CORANE (Swed.)	3/3
	Pronoun	CEDEL2	1/1
		CORANE (Jap.)	6/5
		CORANE (Swed.)	1/1

Table 2: Distribution of SVCs' errors in the three samples

## 5. DISCUSSION

After having reviewed the different errors registered in the three sub-corpora, the discussion will deal fundamentally with three error types: (i) those having to do with the substitution of one or both of the LUs that form the SVC; (ii) those affecting the determiner of the predicate noun; and (iii) analysis errors, which are a recurring phenomenon in only one group of the learners studied: those with Japanese as their L1. The reason for devoting part of the discussion to substitution errors derives from the fact that as well as being the most frequent type of error in the corpora they are also to a certain extent related to the unpredictability of the support verb. As for determiner

errors, they are, along with preposition errors, one of the most frequently repeated kinds of mistake related with the grammatical domain of the error typology applied in this study. However, although it is clear that the selection of a preposition is determined by the particular verb or noun acting as the collocate or the base of an SVC, it is more difficult to establish whether the choice of a determiner has to do with general principles related to noun features (mass noun vs. count noun, identifiable vs. non-identifiable) or with collocational idiosyncrasies. Finally, since analysis errors are repeatedly attested in only one sub-corpus, it is interesting to explore the repercussions of this fact in determining the influence of L1 on collocational mistakes.

### 5.1. Substitution errors

As stated above, substitution errors may affect both the predicate noun and the support verb of an SVC, although most errors affect the former. One main source for this type of lexical mistakes is the transfer of a structure from the L1 to the target language. In fact, this source is the most frequent one for lexical errors in the collocations of CEDEL2 —according to Alonso Ramos et al (2010a: 3212), transfer errors account for more than 70% of collocation lexical errors. As has been already indicated, in the case of the SVC errors of the CORANE corpus no explanation of the annotated mistakes has been attempted. However, it seems plausible to posit the transfer of structures of the L1 or other previously acquired foreign languages as a source of some of the annotated SVC mistakes, and examples such as *ir en huelga* instead of *hacer huelga/ir a la huelga* (‘go on strike’) or *tener posiciones* instead of *tener puestos* (‘to have positions’) in the Swedish-speaker sample seem to point in that direction (cf. Swedish *gå i strejk* or English *position*).

The lack of a directly equivalent SVC in the target language is not the only problem learners have to face when producing an SVC. Another difficulty posed by

SVCs derives from the unpredictability in the choice of the support verb<sup>5</sup>, which can lead to inappropriate use of support verbs that seems to be the result of a learner's hypothesis about the synonymy relations of a given verb or its semantic compatibility with a given noun.

Errors such as *coger el poder* instead of *tomar el poder* ('to seize the power'), or the several incorrect instances of *huelga* ('strike') in SVCs may illustrate the inappropriate use of certain support verbs that in other contexts may be considered synonymous. With regard to *huelga*, this noun has been found as the object of four different verbs (*hacer*, *efectuar*, *realizar* and *practicar*) that can be regarded as synonymous (with the meaning of 'to do, to make'). Only the construction with *hacer*, however, has been considered to be correct by the annotators:

3. las huelgas en España — de que tratan o como y por quienes estan #realizados  
(‘strikes in Spain – what are they about and by whom are they performed’)  
(CORANE C1 092)
4. ¿Por qué los españoles #efectuan huelgas? (‘Why the Spaniards perform strikes?’) (CORANE C1 061)
5. la imposibilidad de #practicar la huelga durante la dictadura que sufrió España  
(‘the impossibility of performing a strike during the dictatorship Spain suffered’)  
(CORANE C1 092)

The decision of marking cases like 3-5 as errors may be a controversial one, since all three combinations can be attested in a large corpus. The four alternatives discussed have been searched and found in the esTenTen11 corpus<sup>6</sup> (Kilgarriff and Renau, 2013). However, the frequency with which the three combinations classified as incorrect are

attested differs substantially from that associated to the default option *hacer huelga*, as the following table shows<sup>7</sup>.

SVC	Frequency per million words	Absolute frequency
<i>hacer huelga</i>	1.2	2,948
<i>realizar una huelga</i>	0.13	339
<i>efectuar una huelga</i>	0.003	8
<i>practicar una huelga</i>	0.002	3

Table 3: Frequency of SVCs with *huelga*

The most frequent combination marked as incorrect is almost ten times less frequent than the default SVC for this meaning and the other two combinations are extremely rare. Thus, although in several contexts the four verbs at issue can be considered synonymous, this synonymy relation does not guarantee its acceptability in a given SVC.

The fact that the selection of a support verb is contingent on the predicate noun it takes as its object does not only interfere with synonymy relations but also with other assumptions regarding the semantic compatibility between a given predicate noun and a given support verb. Thus, for instance, a learner could interpret that nouns such as *conversación* ('conversation') or *costumbre* ('habit'), referring in this particular case to the activity of smoking, as in 7, are things that one makes but does not have, and therefore, that they combine with the verb *hacer* ('to make'), rather than with the stative *tener* ('to have'), —that is, they do not seem to denote a state:



6. el marido como vuelve a casa agotado, no tiene capacidad de \*hacer ni una conversación con ella. ('since the husband comes exhausted back home, he is not even able to have a conversation with her') (CORANE C1 054)
7. Si los fumadores siguen \*haciendo esa mala costumbre ('If smokers keep having that bad habit') (CORANE C1 069)

The correct choice in both cases, however, is *tener*, which is a bit surprising in the case of *conversación*, since its combination with *tener*, a verb usually regarded as stative, enables it to enter contexts normally avoided by stative predicates. Thus, it can combine with the so-called progressive periphrasis, *estar+gerund*, which according to De Miguel Aparicio (1999: 3013), among others, is incompatible with stative verbs<sup>8</sup>.

8. Además, estaba teniendo una conversación muy intensa en ese momento ('Besides, I was having a very intense conversation in that moment') (esTenTen11)

Likewise, *tener una conversación* may occur in another context rejected by states: the simple past combined with an adjunct forcing a punctual interpretation. Since states are not made up by different phases, i.e. they lack an internal development, it is not possible to focus on a given moment of that inexistent development (De Miguel Aparicio, 1999: 3016), which is not the case of *tener una conversación*, as the following example proves:

9. Hace dos días tuve una conversación con mi padre ('Two days ago I had a conversation with my father')

It seems, then, that it is the predicate noun that selects one or another aspectual reading of *tener*<sup>9</sup>, rather than the other way round (cf. the inadequacy of other uses of *tener* in the above contexts: \**Estaba teniendo (un) padre en ese momento*; \**Tuve (un) padre hace dos días*) and that the semantic compatibility between a given predicate noun and a support verb cannot always be predicted.

In sum, substitution errors are not only the outcome of the transfer of learners' L1 structures. In the case of collocates in particular, the fact that their choice is restricted by the identity of the predicate noun adds new difficulties, since synonymy relations do not always help us to select the right support verb nor can its selection always be deduced from its semantic features.

## 5.2. *Wrong determiner*

As the above list shows, together with problematical prepositional choices, one of the most repeated grammatical errors in the samples studied has to do with the use of determiners. As far as prepositions are concerned, it seems clear that their selection is a lexically driven phenomenon, i.e. a certain verb or noun selects a determined preposition or group of prepositions, and it is difficult to make predictions in this respect (for instance, why can we say *tener opción de* and *tener opción a* 'to have the option of', but only *tener la posibilidad de* and not *tener la posibilidad a*, 'to have the opportunity to'?). With determiners, however, it is possible to establish generalizations that predict which determiner will be selected, depending fundamentally on the denotational and referential properties of the base (count vs. mass noun; identifiable vs. non-identifiable). This notwithstanding, it must be taken into account that SVCs add new difficulties, since, as phraseological elements, some of them exhibit several idiosyncrasies in this respect. In this section, two groups of determiner mistakes will be reviewed: (a) those resulting from flouting the patterns governing determiner

distribution in general and (b) those produced by the lack of awareness of the particular behaviour of certain SVCs regarding determination. Since it is hardly predictable when the determiner of an SVC will abide by the general rules of determiner distribution or will follow an idiosyncratic pattern, errors related with the second factor are hard to avoid if the collocation at issue is unknown to the learner.

In non-phraseological combinations, one distinction that is crucial when deciding which determiner to use with a certain noun is the one that opposes count nouns and mass nouns. In general, mass nouns following a verb (i.e., in non-topic position)<sup>10</sup> can be encoded as bare nouns, but count nouns need pluralisation or determination (a quantifier or an article) (cf. Laca, 1999: 894).

10. Falta leche ('Milk is lacking')

11. Como pan ('I eat bread')

12. Como rosquillas/una rosquilla/\*rosquilla ('I eat doughnuts/a  
doughnut/\*doughnut')

In the sample analysed, there are a number of determination errors that consist precisely of choosing a determiner that is either incompatible with a mass-noun reading or produces an undesired interpretation when combined with such a noun. The following examples fit into this picture:

13. Si por ejemplo un político tuviera *\*un sentido de humor* [...] ganaría más votos  
(‘If, for instance, a politician had a sense of humor, he would win more votes’)  
(CORANE C1 090)

14. Me encanta los miércoles en Barnon porque tocan #la música funky. ('I love Wednesday in Barnon because they play the funky music') (CORANE C1 093)

15. [...] si no tendríamos #el humor, [...] ('if we hadn't the humor') (CORANE C1 092)

The quantifier in 13 is incompatible with the non-count character of the phrase *sentido del humor* ('sense of humor'). In 14 and 15, the article induces a universal (or *toto-generic*, in terms of Christophersen, 1939, assumed by Laca, 1990) reading that does not fit in the respective contexts. In 14 the appropriate interpretation would have been a non-specific reading of the phrase *música funky*, since the speaker is not referring to the whole of funky music, but the pieces they play in a pub once a week. The *toto-generic* reading would be acceptable, and the definite article would therefore be necessary, in other contexts such as, for instance, *Me gusta\*(la) música funky* (*I like funky music*). In the other two examples, the determiners would also be acceptable if their contexts are modified. The inclusion of an evaluative adjective in 13 would have rendered the indefinite necessary: *Si un político tuviera \*(un) gran sentido del humor* ('If a politician had a big sense of humour'). Likewise, the inclusion of a restrictive modifier in 15 would require the presence of a definite determiner: *Si no tuviéramos \*(el) humor que se necesita...* ('If we hadn't the humour one needs...') (cf. Leonetti, 1999: 799-800).

Count nouns pose similar problems as far as the election of their determiners is involved. The noun *problema* (problem) occurs three times in three different informants of the Japanese subcorpus as the direct object of the support verb *tener* (have) in the singular and with no determiner, which is ungrammatical, as has already been pointed

out. In order for them to be acceptable, they must have occurred as plural nouns (*tener problemas*) or with an indefinite determiner (*tener un problema*).

As in the preceding cases, there are instances of determiner choice which imply what seem to be unintended readings, as in

16. aunque no diga a su mujer #las palabras de amor (lit. ‘though he doesn’t tell his wife the words of love’) (CORANE C1 054)

As in 14, the selection of the definite article plus a plural count noun forces a *toto-generic* interpretation that seems unintended. The appropriate reading for this context would have been a non-specific one, encoded by means of a plural bare noun: *aunque no diga a su mujer palabras de amor*.

The classification of the next item as an error presents some difficulties. At first sight, *bebida* could be considered a mass noun, so that an example such as 17, with a singular non-count object, would be fine.

17. no tiene que dejarla cuando toma #Ø bebida en un bar (CORANE C1 061)

In Spanish, however, the re-categorization of certain mass nouns as count nouns and vice-versa is not infrequent. Thus, we could say *Yo nunca tomo café* (‘I never drink coffee’) and *Me tomé un café con un amigo* (‘I had a coffee with a friend’), *Me comí un pollo entero* (‘I ate up a whole chicken’) or *Siempre come pollo* (‘He always eats chicken’). The context of 17 seems to require the count version of *bebida*. In fact, examination of a random sample of 100 instances of the verb *tomar* taking *bebida* as a post-verbal complement extracted from the esTenTen11 corpus shows that the noun

occurs always as a count noun (either in plural, or taking a determiner/quantifier, or both).

The preceding mistakes derive from the incompatibility of a particular determiner or its absence with the denotational properties of the base (mass-noun/count-noun; identifiable/non-identifiable). They affect the base of SVCs, but could be solved by the application of general patterns of determiner distribution in Spanish. The following mistakes, in contrast, could be attributed to the idiosyncratic behaviour of SVCs regarding determination patterns. According to what has been said before, object count nouns are not expected to occur in the singular and without a determiner. In the examples below, however, in spite of the fact that the predicate nouns of the SVCs at issue are count nouns, they reject the presence of a determiner.

18. *las he visto yo [las noticias] porque no tengo ??una vida, sabes* ('I have seen the news, because I don't have a life, you know?') (CORANE C1 085)
19. *Y no puede tener \*la vida propia* (And she can't have her own life) (CORANE C1 053)
20. *Personas hacen \*la huelga* ('People go on strike') (CORANE C1 089)
21. *Si no tiene \*el fin, no es viaje* ('If it doesn't have an end, it isn't a trip') (CORANE C1 049)
22. *Tenemos prohibido tener #una pistola.* ('It is forbidden for us to have one gun') (CORANE C1 056)

The examples above seem to fit with two idiosyncratic patterns in SVC determination identified by Alonso Ramos (2004a: 197ff): free null determiners and fixed null determiners<sup>11</sup>. Thus, some of these examples could perhaps be acceptable if

the predicate noun were modified, as in *van a hacer la huelga convocada para el jueves* ('they are joining the strike called for Thursday'). In this same line, both *tener pistola* and *tener una pistola*<sup>12</sup> are possible, but the indefinite determiner induces a precise quantification ('one and only one gun') that would be absent from the undetermined version (*tener pistola*), and the annotators have understood that the latter is the intended interpretation. Others seem to reject determiners even if they are modified (*\*tener la vida propia*).

Different explanations have been proposed for the peculiar pattern of count nouns in this kind of context. Several scholars have related it to their predicative character. Thus Alonso (1933: 137-139) draws a parallel between the facts of a noun's having a predicative meaning and its lack of determiners. His ideas have been taken up again in more recent studies, such as Bosque (1997: 15) or Laca (1999: 894)<sup>13</sup>. Copceag (1964: 197) observes that a sentence such as *Mi amigo tiene coche* ('My friend has a car') does not focus on the possession of a car by a friend, but it characterizes the possessor — Bosque has made similar considerations (Bosque, 1996: 42). The peculiar behaviour of count nouns occurring as objects of verbs like *tener* has also been related with the stereotypical character of the event described by the verb+object combination or with certain cultural expectations (cf. Bosque, 1996: 41ff; Laca 1999: 919). From the learner perspective, however, it is difficult to predict when a situation is stereotypical enough to allow for the absence of determiner in a verb+noun combination. For instance, *padre* ('father') may occur as a singular bare noun as the object of *tener* (*No tengo padre*, 'I don't have a father'), in contrast with *hijo* (*No tengo ??hijo/hijos*; 'I don't have child/children'). It could be argued that *hijo* requires quantification because it is assumed that we have only one father, but can have several children. If that were the case, however, it would be difficult to explain the validity of combinations of *tener* plus bare count nouns

like *tener coche*, *tener perro*, etc. It becomes clear, then, that the idiosyncrasies regarding the use of determiners are a phenomenon related to particular lexical units (cf. Laca 1999: 918) and that attempts at their explanation based on cultural expectations may shed light on the reasons for the lack of determiners, but have little predictive power.

In sum, determination errors in SVC may be the result of disregarding the general patterns of determiner distribution (mass noun vs. count noun, identifiable vs. non-identifiable), but they may also be the product of ignoring certain idiosyncrasies that do not fit into these general patterns. Since the idiosyncratic cases are hard to deduce, they should be treated individually in the learning process.

### 5.3. *Analysis errors*

In the error typology that has been applied to the collocations from the CEDEL2 and CORANE corpora the label *analysis errors* covers those examples where a SCV-like expression has been employed instead of a one-word LU (see above). As already pointed out, this type of error is practically inexistent in samples from the English- and Swedish-speaking learners, but is not uncommon in the sample from Japanese-speaking learners, where 13 instances of analysis have been identified in texts produced by eight different speakers.

Amongst the group of analysis errors, some examples are hard to classify and it could be interpreted that an appropriate version of these expressions could also be an SVC, but with a different support verb, as in the following examples:

23. Si no **\*hagan** estas **soluciones** [solucionan esto/me dan alguna solución], me gustaría cambiar el argumento de contrato [sic] ('If you do not provide any



solution, I would like to change the terms [?] of the contract’) (CORANE C1 050)

24. Por esto, **\*ponían comparaciones** [sic] de [hacían comparaciones de/comparaban] la misma palabra en ambas formas (‘Therefore, they compared the same word in both forms’) (CORANE C1 064)

25. me parece bien que algunos programas **\*ponen sus colaboraciones** [presten su colaboración/colaboren] en las actividades de ayudas (‘I think it is right that some [TV] programmes cooperate with helping [?] activities’) (CORANE C1 064)

In the next example the problem lies in text segmentation: *me hice una caída* could be a substitute for the verbal form *me caí* (I fell) or *una caída* could be not an object of *hice*, but a circumstantial adjunct.

26. Tengo una cicatriz detrás de la cabeza, que **\*me hice una caída** [me caí] cuando era pequeño (‘I have a scar on the back of my head, [because]? I fell [lit. I did myself a fall] when I was a child’/‘I have a scar on the back of my head, that I received [in] a fall when I was a child’) (CORANE C1 042)

The classification of the rest of the examples from this group presented fewer difficulties. Most of them consisted of a predicative noun plus a verb frequently occurring in SVCs, such as *hacer* o *tener*.

27. yo le dije que #**hiciera suficiente preparación** [se preparara] antes de empezar el comercio (‘I told him to prepare himself enough before opening the shop [?]’) (CORANE C1 046)
28. Cuando una mujer puede elegir \***hacer la abnegación femenina** [ser abnegada?] es la virtud (When a woman can choose, to be abnegated is a virtue’) (CORANE C1 053)
29. Desde el septiembre [...] \***haré la actividad profesional** [trabajaré] (‘From September on, I will work’) (CORANE C1 056)
30. Tienen complejo [sic] contra las chicas, no pueden salir ni #**tener comunicación** [comunicarse] con ellas (‘They have a complex about [?] girls, they can’t go out or communicate with them’) (CORANE C1 054)

Some of these SVCs could perhaps be acceptable in different contexts and with slight modifications. Thus, for instance, one can find a considerable number of occurrences of the SVC *desempeñar una actividad profesional* in a Spanish corpus, but it seems that this expression occurs in contexts where certain details (domain, time, conditions) of such “professional activity” are specified:

31. También permite desempeñar la actividad profesional en el ámbito de la enseñanza (‘It also enables to work in the area of education’) (esTenTen11)
32. Los que desempeñen su actividad profesional en el campo de la investigación [...] (‘Those who are engaged in the area of research’) (esTenTen11)

However, using this collocation to convey the meaning of ‘starting a new job’ is highly unidiomatic, as suggested by the fact that in the esTenTen11 corpus (ca. 2.5 billion tokens) the SVC *desempeñar [una] actividad profesional* is attested 226 times,

whereas this same SVC in combination with the inchoative *empezar* is attested only once, and this single instance still includes a reference to professional conditions:

33. [...] donde empezó a desempeñar su actividad profesional como ayudante de cupo del oftalmólogo albaceteño [...] ('where he started to work as assistant to the ophthalmologist from Albacete') (esTenTen11)

In spite of the difficulties in classifying some of these errors, there is a clear contrast between their practical absence in the English- and Swedish-speakers' samples and their recurring presence in the Japanese-speakers' sample, which suggests an influence of the mother tongue on the production of this type of structures. One Japanese light verb in particular has attracted the attention of several researchers in recent decades: the verb *suru* (cf. for instance Grimshaw & Mester, 1988; Miyamoto, 2000; Alonso Ramos, 2001; or Lombardi-Vallauri, 2009). This verb takes as its objects verbal nouns borrowed mainly from Chinese with a remarkable frequency (Miyamoto, 2000: 7; Lombardi-Vallauri, 2009: 183). The bulk of the discussion has revolved around two questions: (i) whether *suru* does in fact have a light version or is simply an ordinary verb capable of assigning thematic roles to their arguments and (ii) if the relation between the verbal noun and *suru* is of a morphological (compositional) or syntactic nature. This, however, is not the focus of interest of the present study, but the fact that the data from the Japanese learners does suggest that its support verbs have an array of combinatorial possibilities that differ from their Spanish counterparts more widely than English or Swedish light verbs do<sup>14</sup>.

## 6. CONCLUDING REMARKS

The examination of the errors affecting SVCs revealed that their production involves a complex array of lexical and grammatical features. This article has focused on two of these in particular: the choice of the support verb (a lexical one) and the selection of the determiner (a grammatical one). With respect to the former, the fact that its choice is mainly conditioned by the identity of a given predicate noun and only to a limited extent is it made on the basis of its weak lexical meaning can cause some otherwise successful strategies used by learners, such as choices based on synonymy relations or other semantic affinities between verbs and nouns, to fail. With regard to the choice of determiners, it has been shown that it does not always follow the general patterns, but is sometimes idiosyncratic.

The two above-mentioned facts have obvious repercussions for the learning process. On one hand, they speak in favour of presenting these constructions in real contexts, rather than as isolated lexical units out of which learners build up SVCs, for instance, by means of concordance-based exercises, in line with proposals such as those of Altenberg & Granger (2001), among others. Learners can thus familiarise themselves not only with the restricted lexical co-occurrence of certain verbs and nouns, but also with their grammatical idiosyncrasies. On the other hand, it seems reasonable that both teachers and materials should pay special attention to those lexical and grammatical aspects of SVCs that are not easily predicted from general patterns. Some materials already do this, for example dictionaries based on the Explanatory and Combinatorial Lexicology model (cf. Alonso Ramos, 2004b, for Spanish), which not only specify the restricted lexical co-occurrence of a given lexical unit, but also features such as determination and prepositional government. Future research focusing on the treatment of these problematic constructions in other learning materials would be desirable.

In addition to the problems with collocate choice and determiner errors, a third type of error has been the subject of special attention due to its recurring character in only one of the sub-corpora: the regular presence of analysis errors in Japanese-speakers' texts. It has been hypothesized that these are the result of a greater productivity of Japanese SVCs as compared to their European counterparts. Such a hypothesis would benefit from further research, as would the impact of different L1s on the collocational performance of learners of Spanish.

#### NOTES

1 They also point out that a significant proportion of the collocations contained in the *Diccionario de colocaciones del español* (Alonso Ramos, 2004b) are SVCs.

2 Common European Framework of Reference for Languages (Council of Europe, 2011).

3 The reader is referred to Alonso Ramos et al. (2010a, b) for a detailed description of the annotation process.

4 Mel'čuk's definition of a surface-syntactic object is as follows: "A main surface-syntactic Object of a lexical unit L is either its direct Object (if L can have one), or its indirect Object (if L cannot have a direct Object), or the strongest prepositional Object (in the absence of the former two types)" (Mel'čuk, 1996: 61).

5 For some researchers, one of the characteristic features of collocates is the unpredictability of their choice, i.e. the fact that they cannot be substituted by any LUs with a similar meaning, their presence being determined by the lexical identity of the base (Mel'čuk, 2012: 39; Nesselhauf, 2003: 224ff). This is, however, a controversial issue in the literature on Spanish collocations. Bosque, taking a diametrically opposed perspective, according to which it is the collocate, as a predicate, that sets the restrictions regarding the arguments it is lexically compatible with, argues that

collocations are not binary relations, but relations that hold between a given predicate and a lexical class of arguments compatible with it (Bosque, 2001, 2011). Thus, collocations can in general be predicted from the meanings of the units combined and unpredictable cases are exceptional. Bosque, nevertheless, acknowledges that in certain collocational domains unpredictability is more widespread than in others, and SVCs are a case in point (Bosque, 2011: xxi). The fact that the choice of support verbs is to some extent unpredictable might be a result of their lexical emptiness: this idea is present, for instance, in Samvelian, Danlos & Sagot (2014), if only to reject the emptiness of light or support verbs and make an attempt to predict their distribution.

6 See <http://www.sketchengine.co.uk/documentation/wiki/TextCorporaOverview>

7 The search has included all the word-forms grouped under the lemma *hacer* with the forms under the lemma *huelga* as post-verbal object.

8 It must be noted, however, that the combination of stative verbs with the progressive periphrasis is possible if these verbs acquire a dynamic nuance, for instance if the initial stage of the state is focused —ingressive interpretation— (de Miguel Aparicio, 1999: 3013).

9 The fact that the object affects the aspectual reading of its verb is not exclusive of support verbs. For instance, depending on the quantificational properties of its object, we can obtain bounded or unbounded event readings of the verb *comer* ('to eat'):

(i) Estaba comiendo carne, pero paró porque le empezó a doler el estómago ->

Comió carne ('He was eating meat, but he stopped because his stomach started hurting -> He ate meat')

(ii) (Se) estaba comiendo un pollo, pero paró porque le empezó a doler el estómago -

> No (se) comió un pollo ('He was eating a chicken, but he stopped because his stomach started hurting' -> He did not eat a chicken)

10 The sequential order seems to be a more relevant feature in this regard than the subject/object distinction. Thus, it is possible to say *Vienen hombres* (lit. ‘come men’), *Pela patatas* (lit. ‘peels potatoes’) but not ??*Hombres vienen* or ??*Patatas pela* (perhaps the latter are acceptable with a contrastive interpretation). In addition to this, it must be noted that non-unaccusative subjects seem to reject bare nouns: cf. *Falta leche* and *Me gusta la/\*∅ música* (‘I like music’).

11 Alonso Ramos (2004a: 198) also cites cases of fixation of the definite article, e.g. *hacer la guerra* (‘to wage war’), *tener la culpa* (‘to be X’s fault’), etc.

12 The inclusion of *tener pistola* as an SVC may be questionable. However, the fact that *pistola* can be regarded as a quasi-predicate and its particular behaviour in relation with determiners when functioning as an object of *tener* speak in favour of their inclusion in the SVC study. In fact, the determined version could be regarded as non-phraseological, as in it *tener* denotes the possession of a given item (or more than one), whereas the undetermined version may be interpreted as a characterisation, meaning ‘to be armed’, without specifying the quantity of arms one possesses.

13 Laca (1999) has a wider conception of the predicative character of nouns than that of, for instance, Polguère. According to her, all the common nouns are semantic predicates: “[...] los sustantivos comunes son, desde el punto de vista semántico, predicados (conceptos generales) que denotan en tanto unidades léxicas, clases de individuos básicos, tipos de material, clases de colecciones o grupos de individuos, o bien reificaciones [...] de propiedades, eventos, estados, etc.” [... common nouns are predicates (general concepts) from a semantic point of view, denoting, as lexical units, classes of basic individuals, types of matter, classes of collections or groups of individuals or reifications of properties, events, states, etc.] (Laca 1999: 894).

14 The translation of some of the *suru* combinations listed in Lombardi-Vallauri (2009) coincide with some of the errors registered in the corpus: for instance *junbi suru* (‘to prepare’) is resembling to the two instances of *hacer preparación* used instead of the verb *preparar*.

## REFERENCES

- Alonso, A. (1933). Estilística y gramática del artículo en español [Stylistic and grammar of the article in Spanish]. In A. Alonso, *Estudios lingüísticos* [Linguistic studies] (pp. 151–194). Madrid, Gredos.
- Alonso Ramos, M. (2001). Constructions à verb support dans les langues SOV [Support-verb constructions in SOV languages]. *Bulletin de la Société de Linguistique de Paris*, XCVI(1), 79–106.
- Alonso Ramos, M. (2004a). *Las Construcciones con Verbo de Apoyo* [Support-verb Constructions]. Madrid, Visor Libros.
- Alonso Ramos, M. (2004b). *Diccionario de colocaciones del español* [Dictionary of collocations of Spanish]. <<http://www.dicesp.com>>
- Alonso Ramos, M., Wanner, L., Vincze, O., Casamayor, G., Veiga, N. V., Suárez, E. M., & González, S. P. (2010a). Towards a Motivated Annotation Schema of Collocation Errors in Learner Corpora Introduction : The Problem. In N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odiijk, S. Piperidis, D. Tapias (Eds.), *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)* (pp. 3209–3214). La Valletta, Language Resources Evaluation.
- Alonso, M.; Wanner, L.; Vázquez, N.; Vincze, O.; Mosqueira, E.; Prieto, S. (2010b): “Tagging collocations for learners”, S. Granger, M. Paquot (eds.), *eLexicography in the 21st century: New Challenges, New Applications. Proceedings of eLex 2009*,



- Cahiers du Cental 7 (pp. 369–374). Louvain-la-Neuve, Presses universitaires de Louvain.
- Altenberg, B., Granger, S. (2001). The Grammatical and Lexical Patterning of MAKE in Native and Non-Native Student Writing. *Applied Linguistics*, 22(2), 173–195.
- Biber, D. (1993). Representativeness in Corpus Design. *Literary and Linguistic Computing*, 8(4), 243–257. doi:10.1093/lc/8.4.243
- Blanco Escoda, X. (2000). Verbos soporte y clases de predicados en español [Support-verbs and predicate classes in Spanish]. *Lingüística Española Actual*, 1(XXII), 99–116.
- Bosque, I. (1996). Por qué determinados sustantivos no son sustantivos determinados. Repaso y balance [Why are certain nouns not determined nouns. Review and balance]. In I. Bosque (Ed.), *El sustantivo sin determinación. Presencia y ausencia de determinante en la lengua española* [The undetermined noun. Presence and absence of determiners in Spanish] (pp. 13–119). Madrid: Visor Libros.
- Bosque, I. (2001). Sobre el concepto de “colocación” y sus límites [On the concept of collocation and its limits]. *Lingüística Española Actual*, 23(1), 9–40.
- Bosque, I. (2011). Deducing collocations. In I. Boguslavsky & L. Wanner (Eds.), *Proceedings of the 5th International Conference on Meaning-Text Theory* (pp. vi–xxiii). Barcelona.
- Buckingham, L. (2008). Spanish verb support constructions from a learner perspective. *Elia: Estudios de Lingüística Inglesa Aplicada*, 151–179.
- Cestero Mancera, A. M., & Penadés Martínez, I. (2009). *Corpus de textos escritos para el análisis de errores de aprendices de E/LE (CORANE)* (Corpus of written texts for learners’ error analysis. CORANE). Alcalá de Henares, Universidad de Alcalá Henares.

- Council of Europe (2011). *Common European Framework of Reference for: Learning, Teaching, Assessment*. Council of Europe.
- <[http://www.coe.int/t/dg4/linguistic/Source/Framework\\_EN.pdf](http://www.coe.int/t/dg4/linguistic/Source/Framework_EN.pdf)>
- Christophersen, P. (1939). *The articles. A study of their theory and use in English*. Copenhagen, Munksgaard.
- Copceag, D. (1964). Un caso de omisión del artículo en rumano y en los idiomas iberorromances [A case of article omission in Romanian and Ibero-romance languages]. *Revue Roumaine de Linguistique*, IX, 195–201.
- García Salido, M. (2014). O uso de construcións con verbos soporte en aprendices de español como lingua estranxeira e falantes nativos [The use of support-verb constructions in learners and native speakers of Spanish]. *Cadernos de fraseoloxía galega*, 16.
- Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A. P. Cowie (Ed.), *Phraseology: Theory, Analysis and Applications* (pp. 145–160). Oxford: Oxford University Press.
- Grimshaw, J., & Mester, A. (1988). Light Verbs and  $\theta$ -Marking. *Linguistic Inquiry*, 19(2), 205–232.
- Gross, G. (1989). *Les constructions converses du français* [Converse constructions in French]. Genève: Droz.
- Gross, M. (1981). Les bases empiriques de la notion de prédicat sémantique [Empirical bases for the notion of semantic predicate]. *Langages*, 15(63), 7–52.
- doi:10.3406/lgge.1981.1875
- Higuera, M. (2006). *Las colocaciones y su enseñanza en la clase de ELE* [Collocations and their teaching in Spanish as a Foreign Language Class]. Madrid, Arco Libros.

- Howarth, P. (1998). The phraseology of learners' academic writing. In *Phraseology: Theory, Analysis, and Applications* (pp. 161–186).
- Jespersen, O. (1946). *A modern English grammar on historical principles. Part VI: Morphology*. London, Allen & Unwin.
- Kilgarriff, A. and I. Renau (2013). esTenTen, a Vast Web Corpus of Peninsular and American Spanish. *Procedia*, 95, 12–19.
- Laca, B. (1990). Generic objects. Some more pieces of the puzzle. *Lingua*, 81, 25–46.
- Laca, B. (1999). Presencia y ausencia de determinante [Presence and absence of determiner]. In I. Bosque & V. Demonte (Eds.), *Gramática descriptiva de la lengua española* [Descriptive Grammar of Spanish] (pp. 891–928). Madrid, Espasa-Calpe.
- Leonetti, M. (1999). El artículo [The article]. In I. Bosque & V. Demonte (Eds.), *Gramática descriptiva de la lengua española* [Descriptive Grammar of Spanish] (pp. 787–890). Madrid, Espasa-Calpe.
- Lombardi-Vallauri, E. (2009). Lexicalization and morphological activation as criteria for Japanese compound verbs. *Rivista Di Linguistica*, 21(1), 181–208.
- Lozano, C., & Mendikoetxea, A. (2013). Learner corpora and second language acquisition. The design and collection of CEDEL2. In A. Íaz-Negrillo, P. Thompson, & N. Ballier (Eds.), *Automatic Treatment and Analysis of Learner Corpus Data* (pp. 65–100). Philadelphia/Amsterdam, John Benjamins.
- Martín Bosque, A. (2006). *¿Fare es hacer?* Colocaciones en los diccionarios monolingües de aprendizaje de ELE [Is fare hacer? Collocations in monolingual dictionaries for the learning of Spanish as a Foreign Language]. In E. Azorín Fernández, D.; Alvarado Ortega, B.; Climent De Benito, J; Guardiola I Savall, M. I.; Lavale Ortiz, R. M.; Marimón Llorca, C.; Martínez Egido, J. J.; Padilla García,

- X. A. ; Provencio Garrigós, H.; Santamaría Pérez, I.; Timofeeva, L.; Toro Lillo (Ed.), *El diccionario como puente entre las lenguas y culturas del mundo. Actas del II Congreso Internacional de Lexicografía Hispánica* (pp. 444–450). Taller digital de la Universidad de Alicante / Biblioteca Virtual Miguel de Cervantes.
- Mel'čuk, I. (1996). Lexical Functions: A Tool for the Description of Lexical Relations in the Lexicon. In L. Wanner (ed.), *Lexical functions in lexicography and Natural Language Processing* (pp. 37–102). Philadelphia/ Amsterdam, John Benjamins.
- Mel'čuk, I. (2004). Verbes supports sans peine [Support-verbs without effort]. *Linguisticæ Investigationes*, 27(2), 203–217.
- Mel'čuk, I. (2012). Phraseology in the language, in the dictionary, and in the computer. *Yearbook of Phraseology*, 3, 31–56. doi:10.1515/phras-2012-0003
- De Miguel Aparicio, E. (1999). El aspecto léxico [Lexical Aspect]. In I. Bosque & V. Demonte (Eds.), *Gramática descriptiva de la lengua española* [Descriptive Grammar of Spanish] (pp. 2977–3060). Madrid: Espasa-Calpe.
- De Miguel-Aparicio, E. (2008). Construcciones con verbos de apoyo en español. De cómo entran los nombres en la órbita de los verbos [Support-verb constructions in Spanish. On how nouns enter verbs' orbit]. In I. Olza Moreno, M. Casado Velarde, & R. González Ruiz (Eds.), *Actas del XXXVII Simposio Internacional de la Sociedad Española de Lingüística (SEL)* (pp. 567–577). Navarra, Universidad de Navarra.
- Mitatou, Z. (2011). Las construcciones del verbo *dar* en la enseñanza del español como lengua extranjera ( E / LE ) [Constructions with the verb *dar* in teaching Spanish as a Foreign Language]. *redELE*, 23.
- Miyamoto, T. (2000). *The Light Verb Construction in Japanese*. Philadelphia/ Amsterdam: John Benjamins Publishing Company. doi:10.1075/la.29

- Molina-Plaza, S., & de Gregorio-Godeo, E. (2010). Stretched verb collocations with give: their use and translation into Spanish using the BNC and CREA corpora. *ReCALL*, 22(02), 191–211. doi:10.1017/S0958344010000078
- Nesselhauf, N. (2003). The Use of Collocations by Advanced Learners of English and Some Implications for Teaching. *Applied Linguistics*, 24(2), 223–242. doi:10.1093/applin/24.2.223
- Nesselhauf, N. (2004). How learner corpus analysis can contribute to language teaching: A study of support verb constructions. In D. S. Guy Aston, Silvia Bernardini (Ed.), *Corpora and language learners* (pp. 109–124). Philadelphia/Amsterdam: John Benjamins.
- Orol González, A., & Alonso Ramos, M. (2013). A Comparative Study of Collocations in a Native Corpus and a Learner Corpus of Spanish. *Procedia - Social and Behavioral Sciences*, 95, 563–570. doi:10.1016/j.sbspro.2013.10.683
- Polguère, A. (2012). Propriétés sémantiques et combinatoires des quasi-prédicats sémantiques [Semantic and combinatorial properties of semantic quasi-predicates]. *Scolia*, 26, 131–152.
- Samvelian, P., Danlos, L. & Sagot, B. (2014). On the predictability of light verbs. In F. Kakoyianni Doa (Ed.), *Penser le Lexique-Grammaire, perspectives actuelles* (pp. 209-221). Paris: Éditions Honoré Champion.
- Sánchez, A., & Cantos-Gomez, P. (1997). Predictability of Word Forms (Types) and Lemmas in Linguistic Corpora. A Case Study Based on the Analysis of the CUMBRE Corpus : An 8-Million-Word Corpus of Contemporary Spanish. *International Journal of Corpus Linguistics*, 2(2), 259–280. doi:10.1075/ijcl.2.2.06san

Sinclair, J. M., & Renouf, A.,(1985). A lexical learning syllabus for language. In C. Ronald & M. McCarthy (Eds.), *Vocabulary and Language Teaching* (pp. 140–160). New York/London, Longman.

Wanner, L., Verlinde, S., & Alonso Ramos, M. (2013). Writing assistants and automatic lexical error correction: word combinatorics. In I. Kosem, J. Kallas, P. Gantar, S. Krek, M. Langemets, & M. Tuulik (Eds.), *Electronic lexicography in the 21st century: thinking outside the paper. Proceedings of the eLex 2013 conference* (pp. 472–487). Ljubljana/Tallinn, Trojina, Institute for Applied Slovene Studies/Eesti Keele Instituut.