
**V+N MISCOLLOCATIONS IN THE WRITTEN PRODUCTION OF
UNIVERSITY LEVEL STUDENTS**

Magdalena Zinkgräf
Universidad Nacional del Comahue, Argentina
maguizinkgraf@jetband.com.ar

The study of collocation – the restricted way in which two or more words combine across texts – has recently gained much relevance. The acquisition of these combinations has proved an important obstacle for non-native speakers of a foreign language as evinced in recorded unusual combinations of words. This paper analyses nonstandard collocations present in the written production of Spanish-speaking university students of English as a foreign language (EFL), in search of recurrent patterns and strategies that may explain these inappropriate collocations as a way to contribute to EFL teaching and learning.

The results show a considerable influence of learners' mother tongue, which induces them to translate literally from L1 into L2 one or both elements in any given collocation. Semantic overlap between appropriate forms and possible synonyms of either the base or the collocate appears to be another factor leading to error. The study also points to some verbs which in this particular context might require consistent and explicit teaching.

Key words: second language acquisition, foreign language teaching, collocations, verbs, nouns, data base, written production

El estudio de la colocación – la combinación restringida de dos o más palabras en diversos textos – ha cobrado mucha relevancia en las últimas décadas. La adquisición de estas combinaciones se convierte en un obstáculo importante para hablantes no-nativos de un idioma extranjero,

quienes combinan palabras de manera poco usual. Este trabajo analiza las desviaciones colocacionales registradas en la producción escrita de alumnos universitarios de inglés como lengua extranjera cuya lengua materna es el español para explorar estructuras recurrentes y las estrategias que las generan, y contribuir, de esta manera, a la enseñanza y el aprendizaje del inglés como lengua extranjera.

Los resultados señalan la importante influencia de la lengua materna, que impulsa a estos alumnos a traducir, en forma directa, uno o ambos elementos en las combinaciones del español al inglés. Otro factor que conduce a errores es la superposición semántica entre la base o el colocado en una colocación y posibles sinónimos. El estudio indica también la necesidad de una instrucción explícita y sistemática de ciertos verbos en este contexto educativo en particular.

Palabras clave: adquisición de una segunda lengua, adquisición de una lengua extranjera, colocaciones, producción escrita, verbos, nombres, base de datos.

1. Introduction

Multiple studies have explored the acquisition of collocations in learners of a foreign language which have proved quite a difficult aspect of the language to learn, as well as a tell-tale sign of a learner's non-native use of language. The following study is based on a manually-compiled corpus of V+N miscollocations gathered from the written production of 102 Spanish-speaking students taking an English language course in Teacher and Translator Training programmes at a university in Argentina. Analysis of this corpus throws light on the essence of these deviations or atypical combinations and the nature of the mistakes they evince.

This paper presents evidence revealing an important proportion of negative transfer from their mother tongue.

2. Collocation

Collocations involve more than random co-occurrence of two or more lexical items across different texts and across different genres, as in *make the bed* and *lay the table*. For reasons of space, not all the different definitions postulated for the term will be explored here; suffice it to say that there are as many as there are authors who have researched the topic. The issue has been studied from multiple theoretical frameworks, the most relevant of which are British Contextualism (Halliday, Sinclair, Hoey, and Phillips), Lexical Semantics (Cruse) and different branches of Lexicography. Many dictionaries have been published which include or deal particularly with the combination of words, and although these will not be commented on individually, it is worth mentioning that both the Spanish and the English lexicographic movements have widely explored and analysed the issue of collocation in specialised monolingual dictionaries, some of which are oriented to foreign language learners of these languages.

With the purpose of restricting the possible interpretations of what a collocation is, and to make clear what is understood by this term herein, a number of characteristics will be presented as essential to the concept. We shall refer to 'collocation' in those cases where

there is an arbitrary and syntagmatic link between at least two lexemes (verbs, nouns, adverbs, adjectives), for example, *commit a crime*, *blissfully unaware*;

there is a constant and recurrent unity observed across multiple texts;

its meaning can be inferred from the elements that compose it, even though one of them may be used in a figurative sense (but not both);

a certain degree of variability (substitution) is possible at one or two points in the collocation (*exert + power/ control/ influence / authority*);

a certain degree of grammatical variability typical of the syntagmatic relation between the component elements is allowed (e.g., *The power exerted by that group was too strong*);

the appearance of other close-class (prepositions, articles, etc.) or open-class words (verbs, nouns, adverbs, adjectives) between the elements in the collocation does not affect the link between them (e.g., *he exerted **all his** authority to...*);

a certain degree of variation related to parts of speech is allowed (if there exists a link such as collocation between a verb and a noun, the collocation is still acceptable if the verb is nominalised, e.g., *an **exertion of power***);

the combination evinces stored knowledge in the memory of native speakers of a language as part of the meaning of any word (that it collocates with others like *x* or *y*);

the expression of ideas and interaction between speakers is aided through pre-selection of lexical alternatives;

the combinations present a certain specificity, i.e., a particular semantic bond between the meanings of the lexical items conveyed through the choice of words (all nouns accompanying the verb *to exert* share a semantic association with the concept of *power*).

These characteristics may help in the recognition of a collocation, typically classified into two broad groups: lexical and grammatical. Those combinations which involve two open-class words correspond to the former, while the latter combine an open-class word and a particle or preposition. This particular study focuses on one instance of lexical collocations, namely those involving a verb and a noun which performs the function of its direct object, or internal argument.

3. Overview of Studies on the Acquisition of Collocation in Foreign Language Settings

Many authors have investigated the puzzling issue of collocation in a foreign language setting, basically aware of the fact that competence in this respect enables learners to perform almost as proficiently as native speakers of the

language (Paquot, 2005). The main conclusions drawn from various studies into the acquisition of English collocations in learners of different L1 backgrounds are presented below¹:

Polish learners of English successfully extended the meaning of L2 collocations on the basis of an equivalent combination in L1, whereas German learners of English made switching mistakes and blends between L1 and L2 (Biskup, 1992).

The wider the meaning scope of a lexical item, the more the interference of L1 in the collocations that lexical item participated in. Similarly, the more synonyms a word had, the more difficulty learners found to produce a restricted collocation (Biskup, 1992).

Advanced EFL learners' collocational competence did not match their general competence in the foreign language; learners found it difficult to paraphrase collocations they could not produce (Bahns & Eldaw, 1993).

Advanced EFL learners were insufficiently aware of the phenomenon of collocation in general (Bahns & Eldaw, 1993; Farghal & Obiedat, 1995).

Farghal and Obiedat (1995) claim that their Arabic learners "tend to respond to the open choice principle (...) rather than the Idiom Principle"² (p. 326), making use of lexical simplification through synonymy, avoidance, transfer and paraphrasing.

The successful use of overlapping collocations (which combine "apparent openness variability- with actual restriction") and the subsequent production of clusters of verb-noun combinations could be a marker of proficiency among native writers and present difficulties to non-native writers (Cowie & Howarth, 1995, p. 84).

Typically, learners of English tended to use the most open or free type of collocations, which presented almost no challenge to their phraseological competence, avoiding the type of language they might consider problematic (Cowie & Howarth, 1995, p. 89).

Learners in Cowie and Howarth's experiment resorted to the extension of a learnt-by-heart restricted collocation used as the basis for new, inappropriate combinations. The process of combining the figurative senses of abstract verbs and nouns in academic writing was perceived as creative by nature, and therefore, productive³, which led to miscollocations.

Non-native French speakers of English tended to resort to some adverbial amplifiers in excess in comparison with native speakers of the language, as a creative strategy to minimise the chances of making mistakes because the appropriate adverbial combinations were not salient to non-native speakers of English, i.e., they did not perceive them as collocations (Granger, 1998).

Non-native speakers of English made use of repetition when they found their collocational knowledge insufficient to express intended meanings (Huang, 2001).

Japanese learners of English turned to a great extent to delexical verbs (*make, do, have, put*), and as their level of general proficiency in the foreign language increased, the nominal phrase that collocated with the selected verb got more complex in terms of modification (Barfield, 2001).

Japanese learners tended to uniformly mistake false verb-noun collocations (experimental distractors) for real collocations (Barfield, 2001).

Freer combinations were easier to complete and translate by Taiwanese learners of English than pure idiomatic expressions (which allow little or no variation), leading to a great number of incorrect variants (Huang, 2001, p. 9).

Distance between L1 and L2 resulted in more incomplete collocational knowledge and negative transfer (Huang, 2001; Mahmoud, 2005).

When there was insufficient knowledge of the verb component of a collocation, recognition of the collocation was impeded (Barfield, 2003),

indicating that verbs play a determinant role in the production of these collocations.

Taiwanese college EFL learners' use of lexical collocations positively correlated with their overall online writing fluency (Hsu, 2007).

Most of these authors have done research into the acquisition of English collocations by learners of different linguistic backgrounds and agree on the typical strategies that may give rise to awkward combinations, namely repetition and reliance on the belief that restricted combinations are freer. Learners may also overgeneralise the meaning scope of those synonymous items participating in a restricted combination or resort to oversimplification and transfer of collocations from L1. However, there is no record of such a study in Romance-language-speaking learners of English. In the following section, an analysis of the production of atypical V+N collocations by Spanish learners of English is presented.

4. The Study

The study consisted of two phases: firstly, the researcher was to create a corpus of Verb-Noun miscollocations recorded in advanced EFL learners' written assignments at an Argentinean university; secondly, the corpus was to be manually analysed in search of typical patterns that would reveal the source of these errors and the unsuccessful, compensatory strategies that these learners make use of when they need to convey meanings that involve collocations.

4.1. Participants, Data-Collection Procedures, and Data Description

Participants in this research involved 102 learners of English taking either the English IV or English V courses during the academic year 2003/ 04 within the Teacher and Translator training programs at the Universidad Nacional del Comahue, Argentina, who agreed to submit their written work for inspection and analysis. The level of proficiency of these learners in the foreign language was assessed through *X_Lex: The Swansea Levels Tests*, a

vocabulary recognition test (Meara, Milton, & Lorenzo-Dus, 2005) that “measures students’ lexical and structural vocabulary knowledge at the most frequent levels in English” (Meara, Milton, & Lorenzo-Dus, 2005, p. 1)⁴. The mean adjusted score obtained was 4188.56 for English IV participants, and 4612.12 for English V learners, respectively (standard deviation: 420.37 in both cases). The first group’s proficiency level is characterized by the program as “competent” and “intermediate level of performance”, and is equalled to that of a CAE course in the test descriptor (UCLES exams) for the 3750-4240 range. The more advanced group of learners in this study is described as having a “very good level of performance” corresponding to CPE (UCLES exams) with scores ranging from 4500 to 4740.

The data were collected from 13 different practical assignments, including reading comprehension tasks, essays and reviews (a total of 270,170 words). Table 1 shows the distribution of these tasks across courses as well as the topics explored in each of them.

Table 1

Data collection instances by task-type and course

Activity Type	Topic	Course
Reading comprehension + essay	Fashion	IV
Reading comprehension + essay	Swimming	IV
Proposal	Menus at the canteen	IV
Reading comprehension + summary	Maxwell & the laws of Electromagnetics	IV
Reading comprehension + essay	Child-care centres	IV
Essay	Television and violence	IV
Reading comprehension + essay	Oestrogen: science and nature	IV

Questions on a novel		Oryx and Crake	IV
Reading comprehension summary	+	Motherkind	V
Report		A festival	V
Reading comprehension summary	+	Murphy's Laws	V
Review		(students' choice)	V
Reading comprehension + essay		The Mayas	V

To build the corpus, all instances of possible inappropriate verb-noun collocations following the basic structure in (1) were transcribed and analysed - the brackets indicate the possible omission of elements.

(1) [(Det^v) + (Adj) + N] + V + (Det) + (Adj) + N

This procedure was followed by the teacher of the course and an external evaluator.

Examples like those in (2) and (3) constituted a large part of the corpus.

(2) [The female body] emitted [reversal clear (sic) signs]

(3) [you] are training [your muscles]

Other grammatical realisations of the same pattern, such as passive forms (4), a nominal phrase modified by a (reduced) relative clause (5), a to-infinitive purpose adverbial phrase containing a transitive verb (6), both gerunds (7) and other nominalisations of transitive verbs accompanied by their direct objects, were also included in the data-base.

- (4) a change of attitude (that) will be held
- (5) the runaways performed in important cities
- (6) in order to achieve their desire
- (7) doing any kind of exercise or sport.

Those cases in which a phrasal verb was used in a transitive way to form a collocation were not included in the database.

As in Nesselhauf's study (2005), these instances of alleged miscollocations were analyzed manually following a set of steps which included consulting each of the available lexicographical volumes specialised on collocations, namely *BBI Dictionary of English Word Combinations* (BBI), *LTP Dictionary of Selected Collocations* (LTP), *Oxford Collocations Dictionary for students of English* (OCDSE) and *Cobuild Concordance and Collocations Sampler* (CCCS)⁶ to assess whether each combination was acceptable. The data on the possible co-occurrence of a verb and a specific noun were recorded on a chart and a criterion was established to determine its inclusion in the corpus, especially when two dictionaries considered it correct and the others did not. If the latter was the case, a British journalist's acceptability-judgement was the determinant.

Through this procedure, a number of collocations were discarded on the basis of positive evidence in favour of these verb-noun units. A corpus of 241 atypical combinations produced by 102 learners - 1,692 words - was obtained.

4.2. Research Questions

The analysis of the corpus took into account the research questions transcribed below:

Which are the typical V+N miscollocations in the written production of these students?

Can they be grouped according to certain verbs, nouns or meanings?

Where does the mistake lie? Is it due to insufficient knowledge of the verb, the noun or both?

What role does the mother tongue play in these irregularities?

The scope of the investigation was delimited to a classification of the miscollocations encountered in the data, the clustering of problematic verbs and nouns, the possible reasons for these unusual combinations and an analysis of the possible impact of the learners' mother tongue on these mistakes.

5. Results and Discussion

5.1. Atypical Combinations per Participant

The average number of miscollocations in the data base (241) produced by number of participants in the study (102) results in the somewhat insignificant coefficient of 2.4 collocations per learner, i.e., each student is responsible for atypically combining a verb and a noun twice in the thirteen practical assignments analysed. This average might look quite promising in terms of the general collocational competence of this particular group of English learners: after all, if any learner miscollocates two certain verbs and nouns within an individual corpus (of assignments submitted) of around 1500 words, in V+N collocation matters, these participants appear to perform almost like native speakers. A more careful and realistic exploration of the data, however, will indicate otherwise.

If instead we analyse how many of the total miscollocations each student has actually produced, we face a more serious and not so encouraging prospect. The odd combinations in the corpus were coded in such a way that a record was kept of which combination had been produced by which student. A statistical test of frequency distribution (Statistix 8.0⁷) to

determine the frequency of occurrence of collocation errors according to which student/s had made them yielded the results presented in Table 2.

Table 2

Number of miscollocations produced by number of students

N° of miscollocations	Participants	
	N°	%
12	1	1.2%
7	2	2.4%
6	4	4.8%
5	5	5.1%
4	10	10.2%
3	12	12.24%
2	23	23.4%
1	45	45.9%

Table 2 indicates that twelve miscollocations were produced by only one participant; two students were responsible for seven atypical combinations each, four learners made 6 mistakes each, and so on. Almost half the population investigated produced one V+N collocational deviation, and almost a quarter produced the ideally estimated 2. In contrast, 36% of students have concentrated 173 instances of incorrect V+N collocations (71%). In summary, half the students have made one or two collocational mistakes, but a sizeable group appears to have had more trouble appropriately combining verbs and their arguments than the rest.

These findings might be interpreted in two possible ways: either at this stage learners show evidence of a relatively weak collocational competence, or these miscollocations are evidence of the risks they take when expressing their ideas in the foreign language, and they will eventually be corrected and made right.

5.2. Frequent Atypical Combinations

The frequency distribution of the whole V+N units in the corpus has thrown light upon the fact that twenty nine (12%) collocations have been repeated as a whole unit, i.e., both the verb and the noun. Nine combinations have been produced twice, whereas three other cases have been repeated three or four times in the data. Table 3 summarises which combinations have been repeated.

Table 3

Most frequent atypical combinations

Twice-produced		Three-time produced	Four-time produced
<i>do mistakes</i>	<i>train muscles</i>	<i>practise exercise/s</i>	<i>make exercise</i>
<i>do strokes</i>	<i>practise activity</i>		<i>do discovery</i>
<i>take consideration</i>	<i>prove experiment</i>		
<i>do swimming</i>	<i>use clothes</i>		
<i>hold war/s</i>			

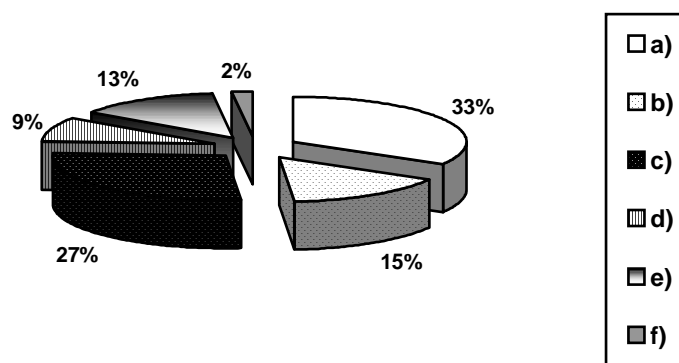
What is striking as regards these miscollocations is that all of them involve very frequent vocabulary items well known to learners in the study, especially considering their advanced level of general proficiency in the foreign language. Most of the verbs involved in these cases are precisely delexical verbs which, on the one hand, are simple and learnt early in the acquisition process, and, on the other, present very restricted collocations; in other words, the typical nouns encountered in the company of *do* will almost never accompany *make*. These very specific restrictions seem to pose a problem to learners, an issue which will be resumed in the following sections.

5.3. Favourite Structures

In section 4.1. the different structural realisations of the V+N relation in question were specified and examples were given of the types included in the corpus. Following Barfield's categories in his study of V+N miscollocations (2003), the data have been analysed in terms of frequency of a certain structure involved in an atypical combination. The categories investigated are:

- a) verb + (det +) noun (e.g., *they are doing a diet*);
- b) noun + verb (e.g., *the runaways performed in important cities*);
- c) verb + (det +) pre-modification + noun (e.g., *they use the same clothes*);
- d) verb + (det +) noun + post-modification (e.g., *you can try to do a lesson of Tae-bo*);
- e) passive form (e.g., *a party was made*);
- f) verb + (det +) pre-modification + noun + post-modification (e.g., *to enrich the intellectual ability of students*)

The results in Graph 1 indicate the percentages of occurrence of these structures in university learners' V+N miscollocations.



Graph 1. V+N miscollocations in terms of structures

The favourite constructions are the canonical word order in structure a) (33%) and the same pattern with the inclusion of an adjective pre-modifying the noun involved in the collocation c) (27%). These two categories resemble d) and f) in word order, though with varying degrees of complexity. What is striking is the fact that categories b) and e) present a non-canonical order of the elements bound by collocation and together add up to 28% of the miscollocations produced. These more complex renderings of the structure V+N may present more problems to learners of a foreign language because, while processing the information they want to convey, they may lose track of the appropriate verb that should accompany the noun they used as head of the construction.

Another interesting issue is the high percentage (57%) obtained in all the categories in which the natural flow of verb + noun is interrupted. Such is the case of not only structures b) and e) but also c), where an adjective might distract learners' attention away from the nominal element central to the collocation (Alonso Ramos, 1994-1995, p. 18). This figure indicates that more than half the miscollocations in the corpus involve a complex realisation of the basic structure which might, on the one hand, speak well of learners' risk-taking in their production when attempting more

elaborate structures, while on the other, point to a possible reason for the formation of these atypical combinations. When the canonical order is inverted or when the sequence is interrupted by a pre-modifier for the noun, learners may be led to choose an incorrect element for the intended collocation, echoing Nesselhauf's (2005) discussion of syntactic pattern and its correlation with learners' difficulties with collocations.

5.4. Most Frequent Verbs

The analysis of the results so far has centred on the frequency and structural characteristics of the combinations as a whole in the corpus. In this section, however, the occurrence of verbs participating in the miscolllocations is analysed.

Through statistical analysis it has been established that delexical verbs have been extensively involved in the atypical combinations of the corpus. More than a quarter of the total collocations in the data (28%) involves one of the following delexical verbs: *do*, *make*, *take*, *have*, *produce* and *hold*, all of which have also been resorted to by Japanese learners of English in the incorrect formation of V+N collocations (Barfield, 2003). In his study Barfield reflects on learners' tendency to use a delexical verb plus a nominalization instead of the more frequent combination found in native speakers' use of a full, lexical verb replacing the odd V+N construction. Examples in the corpus analysed in this paper are:

(8) *his **discoveries** were not recognised at the time he **did them***

(9) *the fabulous **contribution** he **did** to mankind*

These instances show an awkward relationship between verb and noun which could easily be solved if learners had directly resorted to the verbs *discover* and *contribute*.

Further analysis of the frequency distribution of verbs in the corpus evinces a tendency to employ very common verbs in English. Table 4 presents six specific verbs which appear in more than 6 miscolllocations

each, and for each of them, the number of repetitions found in the corpus (Rep) according to the Statistix 8.0 test of frequency distribution, as well as their frequency rating in the General Service List (GSL)⁸ and in the British National Corpus (BNC)⁹ for comparison.

Table 4

Distribution of the most frequent verbs

Verb	Rep	GSL	BNC	Verb	Rep	GSL	BNC	Verb	Rep	GSL	BNC
<i>do</i>	24	98	18	<i>make</i>	19	37	46	<i>Improve</i>	9	604	865
<i>hold</i>	9	172	197	<i>use</i>	7	65	92	<i>Practice</i>	6	-	2658

Make and *do* are by far the most frequent delexical verbs involved in miscollocations, each repeated 24 and 19 times respectively. Many of the “errors” in this database are due to wrong choice of delexical verbs, especially this minimal pair, and could have been avoided if learners had used the other option. Thus, in

(10) [...] *they do a series of changes*

the combination becomes acceptable if *do* is changed for *make*.

What appears both surprising and worrying is that these learners have otherwise reached an advanced level in their general proficiency of the foreign language, as certified by the scores they obtained in *the Swansea Levels Test X-Lex* (Meara et al., 2005). These mistakes, however, point to an important lack in their collocational competence involving mastery of very basic verbs, like *make* and *do*.

Table 5 presents more very frequent verbs, though these have been repeated fewer than five times, as calculated through Statistix 8.0 test of frequency distribution. These have been classified as “considerably frequent” in opposition to the ones above, which participate in more than 6 collocations in the corpus.

Table 5

Verb	Rep	GSL	BNC	Verb	Rep	GSL	BNC	Verb	Rep	GSL	BNC
<i>produce</i>	4	444	332	<i>take</i>	4	60	54	<i>develop</i>	4	130	411
<i>lose</i>	4	417	362	<i>have</i>	3	9	8	<i>enhance</i>	3	-	2666
<i>achieve</i>	3	-	611	<i>prove</i>	3	687	691	<i>Design</i>	3	262	11810
<i>train</i>	3	-	855	<i>face</i>	3	190	599				

A careful look at the frequency rating of the verbs in tables 4 and 5 in *GSL* and *BNC* provides an intriguing insight into the use of these verbs. Most of these cases are not only frequent in the corpus of miscollocations analysed here, but are all the more so in the language, most of them belonging to the 1,000 most frequent words in English. The fact that these very common words in English appear in atypical collocations produced by advanced learners of the foreign language is evidence of the uneven development of learners' general linguistic and collocational competences. As has been stated elsewhere, Bahns and Eldaw (1993), Barfield (2003) and Farghal and Obiedat (1995) have found this to be the case in their investigations. The data in this study indicate that the verbs used are of so basic a nature that it might be assumed that these learners are at an elementary level in terms of their handling of collocation.

On the other hand, the data might also be interpreted as evidence of a strategy learners resort to when faced with the fact that they are not confident in the use of the alternative lexical verb that would otherwise comprise the meanings of both the delexical verb and the intended noun. This particular strategy has been found to be widespread among FL learners across the studies reviewed above, and a very useful one in that these delexical verbs can be typically accompanied by a great - and unrestricted, in students' eyes - number of nouns. Under the assumption that these verbs lack a specific meaning, learners over-generalise and combine them with any noun under the illusion that there is no restriction to the way they can be used. These findings might attest to learners' reliance on Sinclair (1991)'s Open Choice Principle. According to Hunston (2002), reviewing Sinclair's work on this issue,

meaning [of any group or sequence of words] is made either by the phrase as a whole, operating in accordance with the conventional phraseology, or (less often) it is made by the individual words, operating in accordance with grammatical rules (p. 145).

The results of this study seem to show that if Spanish learners' knowledge of English collocations and their inherent restrictions are incomplete, they will construct meaning by adding up meanings of individual words (The Open Choice Principle), ignoring the "whole phrase" principle (The Idiom Principle). In their study of Arabic learners' collocational competence, Farghal and Obiedat (1995) have reached similar conclusions in this respect.

Although interesting insights can be gained through the analysis of noun frequency in the study of miscollocations, and in terms of semantic clustering of both nouns and verbs, for reasons of space these aspects will not be discussed herein (see Zinkgräf, 2005). Instead, the issue of the impact of learners' L1 on these atypical combinations will occupy the following section.

5.5. The Role of the Mother Tongue

To further investigate the motives for the atypical collocations in the corpus, each element in the combinations (verb and noun) was given a translation into the mother tongue by the researcher. In order to ensure the validity of these translations, four external raters whose mother tongue is also Spanish and whose performance in English is quasi native were asked to accept or reject the given translation for the elements in each miscollocation. These raters' assessments resulted in 61% of these atypical V+N collocations being uniformly accepted as generated by negative transfer from L1. The results, which echo Hoey's (2005) predictions for L2 acquisition of collocations, coincide with those obtained by Bahns (1993), Biskup (1992), Bonci (2004), Farghal and Obiedat (1995), Granger (1998), Mahmoud (2005) and Martyńska (2004) as regards the impact of the mother tongue, in this particular case, Spanish, on the production of inappropriate combinations between verbs and their internal arguments in learners of English.

A further step in the exploration of this issue was aimed at determining which constituent is responsible for the deviation in these cases. To this purpose, each of the miscollocations in the data base was provided with an acceptable alternative to express the intended meaning. For instance, for the combination [*extinguish*] + [*disease*] in

(12) [...] *so as to prevent, cure, or extinguish a disease* (Inf. 20)

the dictionaries consulted offer the lemma *ERADICATE* as an acceptable collocate for *DISEASE*.

This procedure was repeated with every miscollocation in the corpus to assess whether the noun, the verb or both in the original version needed to be modified. 70.5% of the atypical combinations in the corpus can be made acceptable by replacing the original verb with one suggested by the lexicographical works consulted. On the other hand, the combinations that could be made right by changing the original *noun* only represent 19%. In 9% of the cases it was hard to determine which element had been selected incorrectly, or what message or idea had been intended by the informants in the first place.

These data indicate that verbs are mainly responsible for the incorrectness in these atypical V+N collocations, while showing that participants in this investigation have insufficient knowledge of the right collocates (verbs) the bases (usually nouns) of the collocations typically take. Once again it is worth noticing that out of the 70.5%, a significant number of wrong verbs belonged to the delexical category, and the “mistake” could be corrected by selecting its counterpart (e.g., *make* vs. *do*).

The results evince an intermediate stage in the acquisition of collocations, since there is evidence that proves these learners lack the linguistic tools to produce restricted collocations, which in turn might lead to heavy reliance on L1 for production of V+N combinations. The participants in this investigation have produced collocations that involve delexical verbs to an unusually important extent as compared to native speakers’ performance, even choosing the wrong one among a set of possibilities. What appears significantly worrying is the fact that the miscollocations

recorded involve ideas, concepts and relations that are of common use in basic exchanges in the foreign language.

6. Conclusions

The analysis of the data base obtained through 13 written assignments submitted by 102 Spanish-speaking learners of English at the Universidad Nacional del Comahue, Argentina, shows the different strategies they typically resort to with respect to their collocational competence. Most of the deviations are the result of negative transfer from their mother tongue, (61%) or, in some other cases, of the exaggerated use of inappropriate delexical verbs. The wrong choice of both nouns and verbs alternatively is recurrent in these atypical collocations, responsible for 70% of the miscollocations in the corpus. Other usual strategies employed are repetition of one same verb combined with nouns which do not collocate, simplification evinced in the reliance on the Open Choice Principle when learners ignore the restrictions some combinations are subject to, and extension by analogy, in those cases where they assume that if a noun shares meaning elements with another, they both collocate with the same verb.

The results also point to recurrent patterns in the structures learners use to convey the verb-noun relation. These conclusions indicate a clear mismatch between learners' general competence in the foreign language and their collocational competence, similar to the results in Schmitt et al. (2004) with respect to a possible correlation between size of the 'individual word lexicon' and the 'formulaic sequence lexicon' (p. 64). Although the data and conclusions herein presented have little potential for generalisation to wider educational contexts, many of the results obtained may be made extensive to similar situations in which advanced learners of a foreign language present limitations as regards the written production of verb-noun collocations.

The insights offered by this study point to an important need for the development of these learners' collocational competence through specific awareness-raising activities as to the restricted nature of the phenomenon in question. Moreover, as some authors suggest (Lewis, 2000; Barfield, 2003), special emphasis should be placed on those areas where L1 and L2 differ in

meanings and collocations. Bearing this linguistic distance in mind should help teachers anticipate possible occasions for negative transfer, and thus select the combinations to be worked on in the English class. In addition, a point should be made of focussing on collocations in foreign language teaching from the very beginning at very early stages in the learning process. Thus, a basic collocational competence could be developed, and on its basis, more restricted and complex meanings built and drawn attention to in subsequent, more advanced courses. Enabling learners to perceive language as being made up of recurrent groups of words should be one of a teacher's most important aims and might eventually help students in the production of the foreign language, leading them to rely more on the Idiom Principle than on the word-by-word construal of meaning, thus resembling a native speaker's natural performance in the use of the language.

Notes

¹ These findings have been arranged following a chronological order.

² Reference to Sinclair (1991)'s Idiom Principle and Open Choice Principle is made in section 4.4.

³ See Sinclair (1991)'s Idiom Principle and Open Choice Principle in section 4.4.

⁴ This computer program provides a vocabulary profile in terms of five different levels of vocabulary knowledge. An adjusted score deduces the number of pseudowords from the total scores, thus taking test-takers' guessing into account.

⁵ 'det' should be understood as 'determiner'

⁶ Consulted at www.collins.co.uk/Corpus/CorpusSearch.aspx

⁷ Computer programme downloadable from <http://undernation.com/crack-step7-micro/crack-Statistix+8.0-serial-keygen-free-full-download-torrent.html>

⁸ The *General Service List* is a list of frequent words in English and can be consulted at <http://jbauman.com/gsl.html>

⁹ A hierarchy of the most frequent lemmas in this corpus can be consulted at <ftp://ftp.itri.bton.ac.uk/bnc/lemma.al>

References

- Alonso Ramos, M. (1994-1995). Hacia una definición del concepto de colocación: de J. R. Firth a I. A. Mel'cuk'. *Revista de Lexicografía*, 1: 9-28.
- Bahns, J. (1993). Lexical collocations: A contrastive view. *ELT Journal* 47(1): 56-63.
- Bahns, J. & Eldaw, M. (1993). Should we teach EFL students collocations? *System* 21(1): 101-114.
- Barfield, A.W. (2001). Collocation design and development. Paper presented at the 11th Vocabulary Acquisition Research Group Network Conference, Swansea University, UK.
- Barfield, A. (2003). Collocation recognition and production: Research insights. Tokyo: Chuo University, Japan.
- Biskup, D. (1992). L1 influence on learners' renderings of English collocations: A Polish/German empirical study. In P.J.L. Arnaud, & H. Béjoint (Eds.), *Vocabulary and applied linguistics* (pp. 85-93). London: Macmillan.
- Bonci, A. (2004). Collocational restrictions in Italian as L2: The relevance of authentic language use for language learning [On-line]. *Arachnofiles*, Vol.3 Hostname: selc.ed.ac.uk
Directory:arachnofiles/pages/three_bonci.htm
- Cowie, A.P. & Howarth, P. (1995). Phraseological competence and written proficiency. In G.M. Blue & R. Mitchell (Eds.), *Language and education. British studies in applied linguistics 11* (pp. 80-93). Clevedon: Multilingual Matters.
- Farghal, M. & Obiedat, H. (1995). Collocations: A neglected variable in EFL. *International Review of Applied Linguistics* 28(4): 315-331.

- Granger, S. (2001). Prefabricated patterns in advanced EFL writing: Collocations and lexical phrases. In A.P. Cowie (Ed.), *Phraseology: Theory, analysis, and applications* (pp. 145-160). Oxford: Clarendon Press.
- Hoey, M. (2005). *Lexical priming. A new theory of words and language*. Oxon: Routledge.
- Hsu, J. (2007). Lexical collocations and their relation to the online writing of Taiwanese college English majors and non-English majors. *Electronic Journal of Foreign Language Teaching* [On-line serial], Vol. 4 (2): 192-209, Directory: e-flt.nus.edu.sg/v4n22007/hsu.htm
- Huang, L. (2001). Knowledge of English collocations: An analysis of Taiwanese EFL learners. *Texas Papers in Foreign Language Education* [On-line serial], Vol. 6, N°1. Hostname: utexas.edu Directory: flesa/tpfle/contents7.doc
- Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.
- Lewis, M. (Ed.). (2000). *Teaching collocation: Further developments in the lexical approach*. Hove: L.T.P.
- Mahmoud, A. (2005). Collocations errors made by Arab learners of English. *Asian EFL Journal* [On-line serial], Vol. 5. Hostname: asian-efl-journal.com Directory: pta.August_05_ma.php
- Martyńska, M. (2004). Do English language learners know collocations? *Investigationes linguisticae* [On-line serial], Vol.XI, Poznań. Hostname: staff.amu.edu.pl Directory: ~inveling/index.php?direct=7
- Meara, P., Milton, J. & Lorenzo-Dus, N. (2005). *The Swansea levels test X_lex, Lognostics CD-Rom* [Computer software]. Express Publishing.
- Nesselhauf, N. (2005). *Collocations in a learner corpus*. Amsterdam: John Benjamins.

- Paquot, M. (2005). EAP vocabulary in learner corpora. A cross-linguistic perspective. [On-line]. In C. Cosme, C. Gouverneur, F. Meunier & M. Paquot (Eds.), *Proceedings of the phraseology 2005 conference*, (pp. 323-326) Hostname: cecl.fltr.ucl.ac.be Directory: Downloads/Paquot%202005.doc
- Sinclair, J. (1991). *Corpus, concordance and collocation*. Oxford: Oxford University Press.
- Schmitt, N., Dörnyei, Z., Adolphs, S. & Durow, V. (2004). Knowledge and acquisition of formulaic sequences. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing and use* (pp. 55-86). Amsterdam: John Benjamins.
- Zinkgräf, M. (2005). Colocaciones atípicas V+N en inglés como lengua extranjera: un análisis de la producción escrita de alumnos universitarios. Unpublished master's thesis, UNED, Madrid.

First version received: March 2008.

Final version accepted: October 2008.