UNIVERSIDAD DE CANTABRIA

FACULTAD DE EDUCACIÓN

MÁSTER UNIVERSITARIO EN APRENDIZAJE Y ENSEÑANZA DE SEGUNDAS LENGUAS / SECOND LANGUAGE LEARNING AND TEACHING

A COMPARATIVE CORPUS-BASED STUDY ON THE USE OF COLLOCATIONS BY NATIVE AND NON-NATIVE SPEAKERS OF ENGLISH

ESTUDIO DE CORPUS COMPARATIVO SOBRE EL USO DE COLOCACIONES EN HABLANTES NATIVOS Y NO NATIVOS DE INGLÉS

AUTOR/A: Andrea Bacigalupi Troule

DIRECTOR/A: Paula Rodríguez Puente

Firma Director/a:

FECHA: 15-09-2015

DECLARACIÓN PERSONAL DE NO PLAGIO

D. / Dña. Andrea Bacigalupi Troule, NIF 72150916P, estudiante del Máster Oficial en Aprendizaje y Enseñanza de Segundas Lenguas / Second Language Learning and Teaching de la Universidad de Cantabria, curso 2014 – 2015, como autora de este documento académico, titulado:

A COMPARATIVE CORPUS-BASED STUDY ON THE USE OF COLLOCATIONS BY NATIVE AND NON-NATIVE SPEAKERS OF ENGLISH / ESTUDIO DE CORPUS COMPARATIVO SOBRE EL USO DE COLOCACIONES EN HABLANTES NATIVOS Y NO NATIVOS DE INGLÉS

Y presentado como Trabajo Fin de Máster, para la obtención del título correspondiente,

DECLARO QUE

este TFM es fruto de mi trabajo personal, que no copio, que no utilizo ideas, formulaciones, citas integrales e ilustraciones diversas, sacadas de cualquier obra, artículo, memoria, etc., (en versión impresa o electrónica), sin mencionar de forma clara y estricta su origen, tanto en el cuerpo del texto como en la bibliografía.

Así mismo, soy plenamente consciente de que el hecho de no respetar estos extremos es objeto de sanciones universitarias y/o de otro orden.

En Santander, a 15 de Septiembre de 2015

Fdo:

AGRADECIMIENTOS

La realización de este trabajo de fin de máster no hubiera sido posible sin el apoyo de mi familia, amigos y profesores.

En primer lugar, me gustaría dar las gracias a mi directora, Paula Rodríguez Puente, por su dedicación y apoyo durante todo el desarrollo del trabajo.

A todos los profesores del Máster en Aprendizaje y Enseñanza de Segundas Lenguas por la formación recibida y a Francisco Gallardo como coordinador del mismo.

A Ignacio Palacios, de la Universidad de Santiago de Compostela, por permitirme usar el SULEC y por facilitarme todos los datos necesarios para la realización del trabajo.

A Javi, por su cariño y apoyo incondicional.

Abstract

Research on second language acquisition has shown the importance of collocational competence in the production of natural and fluent English. Several studies have explored learners' production of collocations showing their problematic nature. However, the collocational knowledge of Spanish learners has not been investigated in much detail so far. This paper presents a corpus-based study on the collocational production of Spanish learners of English at two different levels of proficiency: intermediate and advanced. Verb-noun collocations with three high frequency verbs (*take*, *make* and *have*) were extracted from a learner corpus (SULEC) and compared with data from a native corpus (BNC). The results of the study provide three main findings. Firstly, learners tend to underuse certain collocations found in the native data but, in contrast, they overuse collocations that are not so frequent in the written discourse of native speakers. Secondly, intermediate students produce a higher number of collocations than advanced students but also a higher number of miscollocations. As previous studies show, learners who attempt to produce more collocations are likely to make errors more often. Finally, the present work confirms previous research in showing that L1 influence tends to be the main source of errors in the learners' production of collocations.

Keywords: collocations, learner corpus, proficiency level, verb-noun collocations, native corpus, second language acquisition

Resumen

Las investigaciones llevadas a cabo sobre la adquisición de segundas lenguas muestran la importancia de la competencia en el uso de colocaciones a la hora de producir un inglés natural y fluido. Varios estudios han investigado la producción de colocaciones por parte de aprendices de inglés, mostrando su naturaleza problemática. Sin embargo, los estudios centrados en el uso de este tipo de estructuras por parte de aprendices cuya lengua materna es el español son escasos. Este trabajo presenta un estudio de corpus sobre la producción de colocaciones de aprendices españoles en dos niveles distintos de competencia: intermedio y avanzado. Con este propósito, se extrajeron las colocaciones formadas por tres verbos de alta frecuencia (*take*, *make* and *have*) seguidos de un

sustantivo de un corpus de aprendices (SULEC). Estos resultados se compararon posteriormente con datos de un corpus de hablantes nativos (BNC). Los resultados del estudio señalan tres conclusiones principales. En primer lugar, los aprendices utilizan con menor frecuencia colocaciones que son típicas entre los hablantes nativos pero, por otro lado, tienden a usar en exceso colocaciones que no son tan frecuentes en el discurso escrito de hablantes nativos. En segundo lugar, los estudiantes de nivel intermedio producen un mayor número de colocaciones que los avanzados, pero también un mayor número de colocaciones incorrectas. Tal y como muestran estudios previos, los aprendices que producen más colocaciones tienen una mayor probabilidad de cometer errores. En tercer lugar, la influencia de la lengua materna aparece como la principal fuente de errores en la producción de colocaciones de los aprendices, confirmando así resultados obtenidos en estudios anteriores sobre este tema.

Palabras clave: colocaciones, corpus de aprendices, nivel de competencia, colocaciones verbo-nombre, corpus nativo, adquisición de una segunda lengua.

CONTENTS

List of tables

List of figures

1. Introduction

Formulaic language has become of great importance in language acquisition over the last few years. The emergence of corpus linguistics and corpora studies has demonstrated that native speakers of a language make use of recurrent multiword patterns (Pawley & Syder, 1983; Sinclair, 1991; Wray, 2002). "Multiple word phraseological units" (Schmitt, 2010, p.117) such as idioms, collocations or lexical phrases, constitute a considerable proportion of the language we produce. Therefore, the competent use of formulaic language is considered to have a major role in the production of natural and fluent English (Pawley & Syder, 1983; Nattinger & DeCarrico, 1992; Wray, 2002; Schmitt, 2004).

Computer-aided analysis of large samples of texts allows carrying out studies which were not possible before. One of the concepts that has emerged together with corpus linguistics is the term collocation. Collocational competence is recognized as an important part of vocabulary acquisition (Nattinger & DeCarrico, 1992; Lewis, 2000) since words do not always have independent meanings when they combine. As Bahns (2003) states, over the last few years there has been a "growing awareness of the importance of lexical collocations for vocabulary learning" (p. 56).

An appropriate use of collocations is unmistakably linked to a proficient language use. It is considered as one of the basic ways that distinguishes native speakers of a language from L2 learners. The collocational knowledge of English learners has been the focus of many studies in the area (see, among others, Granger, 1998; Nesselhauf, 2003; Fan, 2009; Li & Schmitt, 2010; Laufer & Waldman, 2011; Granger & Bestgen, 2014). Comparative studies with native and learner corpora have shown that learners often have problems with collocations in their written and oral discourse (Granger, 1998; Howarth, 1998; Nesselhauf, 2003). Many researchers have demonstrated that learners make "overliberal assumptions about the collocational equivalence of semantically similar items" (Wray, 2002, p. 201-202)

According to Siyanova and Schmitt (2008), learners tend to acquire words individually, without taking notice of their immediate environment and that is the reason why they combine words that do not fit together when trying to produce an expression. Pawley and Syder (1983) argue that L2 learners produce utterances that, in

spite of being grammatically correct, are not common in the discourse of native speakers. In the same way, Hill (2000) assumes that the problems learners face with collocations come from their idiosyncratic nature since the elements that form a collocation are highly language-specific and non-natives "can find *eat lunch* or *take lunch* a more obvious choice than *have lunch*" (p. 51).

In sum, comparative corpus-based studies have contributed to the area of second language learning in that they shed light on the problems and difficulties which learners of different mother tongues face when learning a new language. Many authors point out to an L1 influence as the main source of problems when learning an L2 (Bahns & Eldaw, 1993; Bahns, 2003; Nesselhauf, 2003, 2005; Fan, 2009; Laufer & Waldman, 2011).

Although studies analyzing the collocational proficiency of L2 learners are not new, the present study stands out from previous research in that it focuses on Spanish learners of English. As far as I am concerned, Spanish learners' production and knowledge of collocations has not been investigated in much detail so far, with the exception of a few studies (Zingraf, 2008; Marco, 2011). Therefore, the aim of the present work is to examine the collocational knowledge of Spanish learners of English by means of a comparison with data extracted from a native corpus. Moreover, it also investigates the relationship between the level of competence of learners and their knowledge of collocations. Finally, the main errors in the collocational production of Spanish learners will be analyzed in order to determine the main sources that account for these errors.

The present work opens with a review of the previous literature on the subject in order to clarify the meaning of the term collocation (Section 2.1), the types of collocations that some authors establish in their classifications (Section 2.2), the criteria established to distinguish collocations from other multiword units (Section 2.3), the previous studies carried out about this topic (Section 2.4) and, finally, a description of the present study (Section 2.5). Chapter 3 describes the methodology of the study, including the corpora used and the techniques to collect and analyze the data. The results obtained are shown and discussed in Chapter 4 in order to answer the research questions of the study. Finally, Chapter 5 summarizes the main conclusions reached.

2. Review of the literature

2.1. Defining collocations

The term collocation is a very controversial concept. Previous literature on the subject reveals a conflict regarding the meaning of this term since it can be used and understood in many different ways (Bahns, 1993). Among all the definitions of the term, Nesselhauf (2004) distinguishes two main views: the frequency-based approach and the phraseological approach.

2.1.1. The frequency-based approach

This approach is related to statistic criteria and understands the term collocation as a frequent co-occurrence of words at a certain distance (Nesselhauf, 2004). J. R. Firth was one of the first researchers who adopted the frequency-based approach. According to his view, "you shall judge a word by the company it keeps" (Firth, 1957, p. 179). This method was further developed by J. Sinclair, who gave a clearer definition of collocation as "the occurrence of two or more words within a short space of each other in a text" (Sinclair, 1991, p. 170). In the same way, Halliday and Hassan (2001) argued that a collocation is "the co-occurrence of lexical items that are in some way or other typically associated with one another, because they tend to occur in similar environments" (p. 317). A similar definition of the term is also given by Lewis (2000) as "the way in which words co-occur in natural text in statistically significant ways" (p. 132).

2.1.2. The phraseological approach

From the perspective of the phraseological approach, collocations are conceived of as a type of word combination fixed to some extent (Nesselhauf, 2004). The main difference between the two approaches is that in the phraseological approach the elements that make up a collocation must be related syntactically (Kurosaki, 2013).

A. P. Cowie is the most influential representative of this approach. He defined collocations in order to delimit them from other word combinations, namely free combinations and idioms (Cowie, 1981). Specifically, he distinguishes four types of word combinations (Nesselhauf, 2005, p.14):

1) Free combinations (e.g. *drink tea*), in which all the elements of the word combination have a literal sense and the restrictions on substitution are based on semantic grounds.
2) Restricted collocations (e.g. *perform a task*), where at least one of the elements has a non-literal meaning and some substitution is possible (*perform a task/function/duty* but not *perform an assignment*).
3) Figurative idioms (e.g. *do a U-turn* 'the turning of a vehicle in a U-shaped course so as to face in the opposite direction' or 'a change of plan, especially a reversal of political policy'), in which the combination has a figurative meaning but retains a literal interpretation. Substitutions are rarely possible.
4) Pure idioms (e.g. *flying colors*), where the combination has a figurative meaning that does not have a literal interpretation; in other words, it is impossible to deduce its meaning from the individual parts. Substitution is not possible.

These four types of word combinations form Cowie's phraseological continuum. As shown in Figure 1, the most variable and transparent elements are on the left hand side, whereas the most opaque and fixed elements appear on the right hand side.

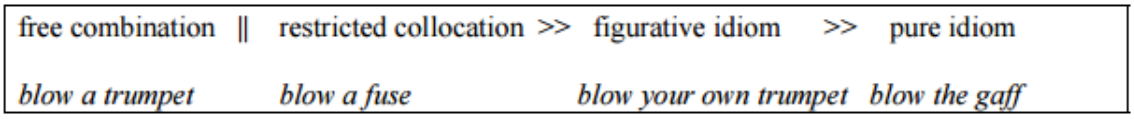

Figure 1: Cowie's phraseological continuum. Reprinted from "*The treatment of collocations and idioms in learners' dictionaries",* by A. P. Cowie, 1981, *Applied Linguistics, 2.*

2.2. Types of collocations

According to Benson, Benson, and Ilson (1986), collocations can be divided into two major groups: grammatical collocations and lexical collocations.

Grammatical collocations are made up of a dominant word, such as a noun, verb or adjective, and a preposition or grammatical structure like a clause or infinitive. Within the grammatical collocations we can distinguish eight major types, as shown in Table 1 below.

Grammatical collocations		
G1	Noun + preposition (Except noun + *of*; noun + *by*)	*Blockade against*
G2	Noun + *to* + infinitive	*They made an attempt to do it.*
G3	Noun + *that* clause	*We reached an agreement that she would represent us in court.*
G4	Preposition + noun	*By accident*
G5	Adjective + preposition	*Angry at everyone*
G6	Predicate adjectives + *to* + infinitive	*It was necessary to work.*
G7	Adjective + *that* clause	*It was nice that she was able to come home.*
G8	Verb + direct object + *to* + indirect object = Verb + Indirect Object + Direct Object	*She sent the book to him =* *She sent him the book.*
	Verb + direct object + *to* + indirect object	*They described the book to her.*
	Verb + direct object + *for* + indirect object = Verb + indirect object + direct object	*She bought a shirt for him =* *She bought him a shirt.*
	Verb + preposition + object	*They came by train.*
	Verb + to + infinitive	*He decided to come.*
	Verb + bare infinitive	*We must work.*
	Verb + V-*ing*	*The house needs painting.*
	Verb + object + *to* infinitive	*He invited me to participate.*
	Verb + object + infinitive	*She heard them leave.*
	Verb + object + V-*ing*	*He kept me waiting two hours.*
	Verb + a possessive + V-*ing*	*They love his clowning.*
	Verb + *that* clause	*They admitted that they were wrong.*
	Verb + object + *to be* + complement	*We consider her to be very capable.*
	Verb + object + complement	*She dyed her hair red.*
	Verb + object1 + object2	*The teacher asked the pupil a question.*
	Verb + (object) + adverbial	*He carried himself well.*
	Verb + (object) + *wh*-clause/*wh*-phrase	*He asked how to do it.*
	It + verb + object + *to* infinitive *It* + verb + object + *that*-clause	*It behoves you to study more.* *It surprised me that our offer was rejected.*
	Verb + Complement (adjective or noun) Verb + Complement (adjective)	*He was a teacher.* *The flowers smell nice.*

Table 1: Grammatical collocations. Adapted from *The BBI combinatory dictionary of English: A guide to word combinations* (p. xvi-xxviii), by Benson et al., 1986, Amsterdam: John Benjamins.

Lexical collocations, contrary to grammatical collocations, do not include prepositions, clauses or infinitives. In general, lexical collocations are made up of nouns, adjectives, verbs and adverbs. There are seven major types of lexical collocations, as shown in Table 2 below.

Lexical collocations		
L1	Verb (denoting creation and/or activation) + noun/pronoun	*Make an impression* *Reach a verdict*
L2	Verb (meaning eradication or nullification) + noun	*Break a code* *Annul a marriage*
L3	Adjective + noun	*Weak tea* *Formidable challenge*
L4	Noun + verb	*Blood circulate* *Alarms go off*
L5	Noun 1 *of* noun 2	*A bouquet of flowers* *A piece of advice*
L6	Adverb + adjective	*Strictly accurate* *Keenly aware*
L7	Verb + adverb	*Affect deeply* *Appreciate sincerely*

Table 2: Lexical collocations. Adapted from *The BBI combinatory dictionary of English: A guide to word combinations* (p. xxx-xxxiii), by Benson et al., 1986, Amsterdam: John Benjamins.

Lewis (2000), in turn, elaborates a list including different types of collocations. Lewis' list is more comprehensive than Benson et al.'s in that it includes a wider number of items, as shown in Table 3 below.

1. Adjective + noun	*A difficult decision*
2. Verb + noun	*Make a mistake*
3. Noun + noun	*Radio station*
4. Verb + adverb	*Examine thoroughly*
5. Adverb + adjective	*Extremely inconvenient*
6. Verb + adjective + noun	*Revise the original plan*
7. Noun + verb	*The fog closed in.*
8. Discourse marker	*To put it another way*
9. Multi-word prepositional phrase	*A few years ago*
10. Phrasal verb	*Turn in*
11. Adjective + preposition	*Aware of*
12. Compound noun	*Fire escape*
13. Binomial	*Backwards and forwards*
14. Trinomial	*Hook, line and sinker*
15. Fixed phrase	*On the other hand*
16. Incomplete fixed phrase	*A sort of...*
17. Fixed expression	*Not half!*
18. Semi-fixed expression	*See you later*
19. Part of a proverb	*Too many cooks...*
20. Part of a quotation	*To be or not to be...*

Table 3: Lewis' classification of collocations. Adapted from *Teaching collocation: Further developments in the lexical approach* (p. 133-134), by M. Lewis, 2000, Hove: Language Teaching Publications.

2.3. Distinguishing criteria

Many linguists have established different criteria in order to make a distinction between free combinations, collocations and idioms. Aisenstadt (1981) differentiates the three concepts in terms of collocational restriction, categorizing English word combinations into idiomatic and non-idiomatic, as seen in Figure 2 below.

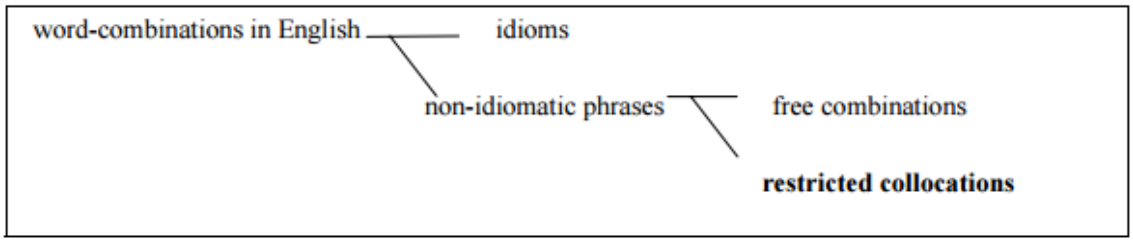

Figure 2: Aisenstadt's categorization of English word combinations. Reprinted from "Restricted collocations in English lexicology and lexicography", by E. Aisenstadt, 1981, *Review of Applied Linguistics, 53,* (p. 54).

Idiomatic combinations are fixed combinations whose meanings do not reflect the meanings of their individual parts (e.g. *red tape* 'bureaucracy'). In turn, non-idiomatic combinations can be divided into, on the one hand, free collocations, which are combinations of two or more words with free commutability and, on the other hand, restricted collocations. The latter are defined as "a type of word combination consisting of two or more words, unidiomatic in meaning, following certain structural patterns, restricted in commutability not only by semantics, but also by usage" (Aisenstadt, 1981, p. 54).

Nattinger and DeCarrico (1992) established three criteria to distinguish idioms, collocations and free combinations, namely flexibility, compositionality and productivity. Based on these parameters they set up a continuum of word combinations, as shown in Figure 3 below.

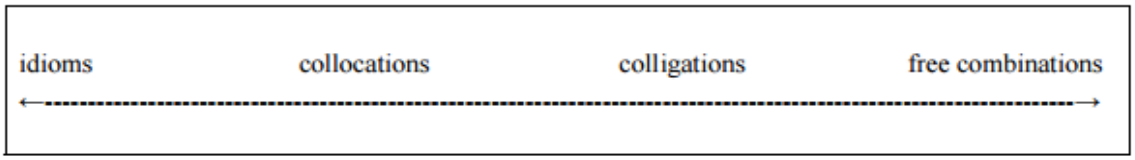

Figure 3: Nattinger & DeCarrico's continuum of word combinations. Reprinted from *Lexical phrases and language teaching* (p. 178), by J. R. Nattinger &J. S. DeCarrico, 1992, Oxford University Press.

At the left-most end of the continuum we find idioms, which are considered as non-compositional, as their meaning is unpredictable, and non-productive, since there are relatively few. At the opposite end we see free combinations, which are fully compositional, completely predictable and productive. In the middle of the continuum we find collocations, which are compositional since they are roughly predictable but restricted to certain specified items.

Carter (1998, p.70) also established a cline in collocational restriction that runs from the least fixed combinations (a) to the most fixed ones (d), as shown below.

(a) **Unrestricted collocation**: Lexical items which are open to partnership with a wide range of items. Most core words and structures with core verbs fall into such category (e.g. *take a look/a holiday/a rest/a walk*).

(b) **Semi-restricted collocation**: Lexical patterns in which the number of items that can be substituted is more limited (e.g. *harbour doubt/grudges/uncertainty/suspicion*).
(c) **Familiar collocation**: Words which keep regular company with each other (e.g. *amicable divorce* or *unrequited love*).
(d) **Restricted collocation**: Partnerships are more fixed and close (*e.g. stark naked* or *pitch black*).

Similarly, Hill (2000, p.63-64) differentiates four types of collocations in terms of collocational strength.

1. **Unique collocations**: This type of collocations is unique, strong and fixed. For example, *foot* is only used as a verb in the collocation *foot the bill* 'to pay for something, especially something expensive'.
2. **Strong collocations**: Although not unique, these collocations are strong or very strong since the number of collocates is limited. We often have *ulterior motives* or *harbour grudges* and the knowledge of the words *motives* or *grudges* would be incomplete without the knowledge of those strong collocates.
3. **Weak collocations**: They can be easily predicted. Many things can be *long* or *short*, *good* or *bad*. Those adjectives can be applied to almost anything, such as *good meal*, *good journey* and *good time*.
4. **Medium-strength collocations**: This category includes all those collocations in between the weak and the strong ones. For example, *hold a conversation* or *make a mistake.*

Another classification is the one established by Conzett (2000) who argues that collocations can be **strong**, if the presence of one word predicts the occurrence of another one, or **weak**, when there is a great variance in the number of words that may collocate with another. She establishes a continuum running from units made of freely-combining words, such as *friendly dog* or *old car* to fixed expressions and idioms like *throw in the towel* 'give up'. The expressions in the middle of the continuum, between free combinations and idioms, are considered as collocations, with the strongest ones to the right and the weakest ones to the left, as shown in Figure 4 below.

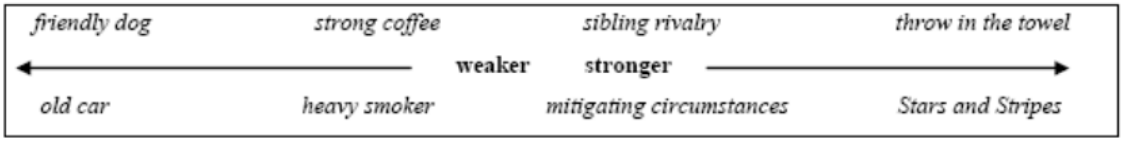

Figure 4: Conzett's continuum model. Reprinted from *Teaching collocation: Further developments in the lexical approach* (p. 74), by M. Lewis, 2000, Hove: Language Teaching Publications.

As we can observe from the literature on the subject, many linguists agree in the existence of a continuum of word combinations. Collocations are always placed in the middle of that continuum, between free combinations and idioms. Most authors establish the same criteria to differentiate between these terms, their degree of restriction and their semantic opacity. However, there seems to be no definite and clear limit between the three terms.

2.4. Collocations and L2 learning

Research on second language acquisition has shown the importance of phraseological competence in order to achieve native-like proficiency. The use of prefabricated patterns is an essential component of second language acquisition and one of the aspects that clearly differentiates native speakers of a language from L2 learners (Granger & Bestgen, 2014).

According to Hill (2000), the lack of collocational knowledge leads L2 students to make grammatical mistakes since "students tend to create longer utterances because they do not know collocations which express precisely what they want to say" (p. 49). One of the main reasons that account for this problem is that "collocations are arbitrary and unpredictable" (Benson et al., 1985, p. 285). Whereas native speakers (henceforward NS) acquire collocational knowledge unconsciously while they grow up in their language community, most non-native speakers (henceforward NNS) do not have that opportunity (Fan, 2009).

The rise of corpus linguistics and the introduction of corpus analysis techniques have allowed researchers to carry out studies involving a large amount of data and cross-corpora analyses. Many of them compare the production of collocations by NS

and NNS of English (Bahns & Eldaw, 1993; Howarth, 1996; Granger, 1998; Nesselhauf, 2003; Fan, 2009) showing the learners' deficiency in collocational knowledge. The number of collocations produced is often inferior in the case of NNS in comparison with NS (Laufer & Waldman, 2011). Results also point out to an underuse of certain combinations found in native data (Granger, 1998) and an overuse of some collocations in situations where more specific meanings are required (Shih, 2000).

There is a wide range of studies related to collocational knowledge and production. The focus of most investigations is on lexical collocations made up of nouns, adjectives, verbs and adverbs (e.g. *deeply absorbed, a bouquet of flowers*) and grammatical collocations formed by a noun, adjective or verb plus a particle (e.g. *in advance, apathy towards*). Generally, research on L2 collocations has focused on one particular structure or type of collocation. For example, Granger (1998) investigates collocations made up of *–ly* intensifiers and adjectives, a type of lexical collocation formed by an adverb and an adjective (see Table 2 above) through the comparison of essays written by native and advanced non-native speakers whose L1 is French. The results demonstrate that learners' phraseological skills are limited. Although they do employ collocations, they use few native-like expressions and tend to include atypical word combinations.

Many researchers have focused on lexical adjective-noun combinations (Siyanova & Schmitt, 2008; Li & Schmitt, 2010; Lukač & Takač, 2013). Siyanova and Schmitt (2008) presented a series of studies focused on L2 production of adjective-noun collocations. The results of the studies suggest that, although L2 learners are capable of producing appropriate collocations, their fluency and underlying intuitions with collocations are not comparable with those of native speakers.

However, one of the most recurrent combinations when investigating the collocational competence of L2 learners is the verb-noun construct (Bahns & Eldaw, 1993; Howarth, 1996; Barfield, 2003; Nesselhauf, 2003; Gyllstad, 2007; Laufer & Waldman, 2011).

Nesselhauf (2003) carried out an exploratory study that examined the use of verb-noun collocations by German-speaking learners of English in written essays. She distinguished three types of collocations in the study: free combinations, collocations and idioms (see section 2.3). The findings report that the higher rate of mistakes (79%) occurs in collocations with a medium degree of restriction, in which the sense of the

noun is used in an unrestricted way but the sense of the verb is restricted (e.g. *take a picture/photograph*; but not *take a film/movie)*, due to the wrong choice of verb. The results also show that L1 influence plays an important role on the production of wrong collocations.

Howarth (1996) examined the use of prefabricated patterns in the academic writing of native and non-native speakers of English. He analyzed a native corpus in order to find out to what extent collocations occur in formal written English. Then he carried out the same analysis in a corpus of advanced learners' writing. The findings show that non-native speakers' mistakes in the use of collocations do not affect intelligibility; however, they may affect the precision and clarity required in academic communication.

Bahns and Eldaw (1993) focused their research on German advanced EFL students' productive knowledge of verb-noun collocations by means of a translation task and a cloze task. They found that collocations constitute a major problem in the students' production of correct English, particularly when trying to paraphrase collocations in order to avoid the difficulties they present for learners since they are idiomatic. They also found out that collocational knowledge does not develop in parallel with vocabulary knowledge.

Laufer and Waldman (2011) also investigated the use of verb-noun collocations in the writing of NS of Hebrew at three proficiency levels. They compiled a learner corpus and compared it with a corpus of native speakers of English (Louvain Corpus of Native English Essays, LOCNESS). The findings show a lower production of verb-noun collocations in the case of L2 learners in the three levels of proficiency. L1 interference also constitutes an important factor in the production of collocations since half of the errors at all levels of proficiency are caused by L1 transfer.

Focusing on Spanish learners of English as a foreign language, Zingraf (2008) gathered a corpus of verb-noun miscollocations ('wrong collocations' e.g. *make a photo* instead of *take a photo*) from the written production of 102 Spanish-speaking students in order to establish the cause of their mistakes. Most of them are the result of a negative transfer from their L1 (61%), while in other cases it was because of the exaggerated use of inappropriate delexical verbs, such as *give, have* or *make*. The wrong choice of verbs and nouns is responsible for 70% of the miscollocations.

Marco (2011) also explored atypical verb-noun collocations in a corpus of English technical writing of Spanish students in an attempt to explore the reasons why learners deviate from NS's norms. The analysis showed that students are unaware of collocations that are typical of technical writing, which suggests that students' knowledge of the specific vocabulary of a discipline may be incomplete since they lack collocational knowledge. The study also points out to L1 transfer as the main source of mistakes in the production of collocations.

2.5. The study

As shown in the review of the literature, there is a high percentage of studies exploring learners' production and knowledge of verb-noun collocations. However, Spanish learners' difficulties with collocations have not been investigated in much detail so far. There is a gap in the literature concerning this type of research and, for that reason, this study attempts to discover the state-of-the-art in the collocational competence of Spanish learners of English by exploring a learner corpus, the Santiago University Learner Corpus (SULEC), and comparing the data obtained with material extracted from a native corpus, the British National Corpus (BNC).

Moreover, the study attempts to explore the relationship between the proficiency level of the learners and their collocational knowledge, given that studies examining the development of phraseological competence as a function of degree of proficiency in L2 are scarce (Granger & Bestgen, 2014).

From the perspective of the phraseological approach, in the present study collocations are understood as a type of word combination fixed to some extent, in which at least one of the elements has a non-literal meaning (Nesselhauf, 2004). However, this work adopts a perspective in which the frequency-based approach and the phraseological approach are not mutually exclusive, since the frequency of certain collocations will also be explored in both corpora.

The study focuses on the written production of one type of lexical collocations based on Benson et al.'s (1986) classification (see Table 2 above), namely the verb + noun structure. It is concerned with the production of restricted collocations made up of

a high frequency verb and a noun, such as for example *make a mistake* or *have a coffee*. The verbs selected for the study are *take*, *make* and *have* since these three verbs are especially productive in combining with noun phrases to form relatively idiomatic expressions (Biber, Johansson, Leech, Conrad & Finegan, 1999, p. 1026).

Spanish learners' production of this type of collocations will be analyzed and compared with data from a corpus of native speakers of English. The research questions addressed in this study are the following:

a) What are the main verb-noun collocations produced by Spanish learners of English and native speakers?
b) Are there any differences in the written production of verb-noun collocations by Spanish learners of English and native speakers?
c) What is the relationship between the proficiency level of Spanish learners and their knowledge of collocations?
d) Which are the main sources of errors in the collocational production of Spanish learners of English? Do they differ from those of learners whose L1 is other than Spanish?

In the following sections, we present the results obtained in order to give answer to these issues.

3. Methodology

3.1. The corpora

Collocations for the study were extracted from two corpora: one of Spanish learners of English and the other of NS. The learner corpus selected for this study is the Santiago University Learner Corpus (SULEC) from the University of Santiago de Compostela, in Spain. On the other hand, the British National Corpus (BNC) has been used in order to compare the results from the learner corpus with real native data. Table 4 presents detailed information about the size and structure of the corpora.

	SULEC	BNC
Total corpus size	480,000 words	100,000,000 words
Oral data	30,000 words	10,000,000 words
Written data	450,000 words	90,000,000 words

Table 4: Size and structure of the corpora.

3.1.1. SULEC

As shown in Table 4 above, the SULEC corpus contains about 480,000 words including both oral and written samples produced by secondary and university students. The spoken data includes 30,000 words and it has been collected through semistructured interviews,[1] short oral presentations and brief story descriptions.

On the other hand, the written data comprises 450,000 words and has been gathered from compositions or argumentative essays. Students were asked to write a 500-word composition about one of the topics given. The subjects of the compositions varied from university to secondary school students to suit learners' interests. Among the topics selected the following were included:

1) The problem of the Prestige.
2) What is your position about the University Entrance Examination?
3) Most university degrees are theoretical and do not prepare students for the real world. They are therefore of very little value.

[1] This means that the interviewers will have an interview outline in front of them. However, this will be quite flexible so as not to condition the informant excessively.

4) Do you think the marriage of persons of the same sex should be allowed?

The samples are divided into three levels of linguistic competence: elementary, intermediate and advanced. These levels have been established depending on the age of the participants. The elementary level includes students from the 3rd year of Secondary Education to the 2nd year of Bachillerato. The intermediate level (240,000 words) comprises students from the 1st and 2nd year of an English Philology degree while in the advanced level (210,000 words) we find students from the 3rd and 4th year of the aforementioned university degree.

3.1.2. BNC

The BNC is a 100 million word collection of samples of written and spoken language from a wide range of sources. The spoken component of the BNC constitutes approximately 10 per cent (10 million words) of the total and comprises transcriptions of informal conversations and spoken language from different contexts, such as formal business or government meetings.

The written section, which constitutes 90 per cent (90 million words) of the corpus, includes newspaper extracts, specialist journals, academic books and university essays, among others. The corpus further includes a wide variety of genres or text types, namely fiction, magazine, newspaper, non-academic, academic and miscellaneous. The academic sub-section, which is the focus of the study, is the most extensive one comprising 15,331,668 words.

3.2. Data collection

In order to answer the research questions addressed in the study, data was collected from the two corpora mentioned above. Regarding the learner corpus (SULEC), verb-noun collocations were selected manually. The process followed three steps:

1. Each of the three verbs was introduced individually in the online corpus browser in all its possible forms (simple form, *-s* form, past tense form, *-ing* form and past participle form).
2. The occurrences of the different forms of the verbs were checked manually in its context, extracting verb-noun combinations and eliminating other cases that were irrelevant for the study (e.g. extracts where the verb is not followed by a noun or verbs used as auxiliaries).
3. The verb-noun combinations extracted were divided into three levels according to the linguistic competence of the students (elementary, intermediate and advanced).

The collection of data in the BNC was a simpler process since the corpus offers online services with more advanced search functions than the learner corpus. The Brigham Young University browser allows looking for a word and its most frequent collocates within the BNC corpus. The browser includes two different fields: one for the main word(s) and the other for the collocates. In addition, it allows looking for the overall frequency of a word in the whole corpus or in a specific section. The process was the following:

1. All the possible forms of the verbs (simple form, *-s* form, past tense form, *-ing* form and past participle form) were introduced in the field for words.
2. In the field for collocates I introduced the code [nn*], which indicates the word class, in this case nouns.
3. I selected the written academic section so as to get comparable data.
4. The 20 most frequent verb-noun combinations in the corpus were extracted.

3.3. Data analysis

The total number of verb-noun collocations extracted from the SULEC was revised individually in order to detect miscollocations. The wrong verb-noun combinations were removed from the list and were analyzed later in order to ascertain the main source of learners' collocational errors.

The remaining verb-noun combinations differed considerably in terms of idiomaticity. According to Biber et al. (1999, p-1026-1027), the resultant expressions from the combinations of the verbs *take, make* and *have* and a noun form a cline of idiomaticity. The cline runs from idiomatic expressions, such as *have a look* or *take time,* to expressions that retain the core meaning of the verbs, such as *make a sandwich* or *take a snack.* In between we can find semi-idiomatic expressions, in which the meanings of individual words are retained to some extent, but the whole expression has a more idiomatic meaning. Some examples of these expressions are *have dinner* or *take part.* In addition, many of these expressions can be replaced by just one verb (e.g. *have dinner – dine*).

Since the classification of verb-noun combinations into a cline of idiomaticity goes beyond the scope of this study, the following criterion has been established in order to classify the total number of verb-noun combinations. Taking into account that at least one of the elements that form a restricted collocation has a non-literal meaning, all the expressions in which the verb retains its primary meaning have been removed.

According to Hornby (2000), editor of the *Oxford Advanced Learner's Dictionary of Current English* (OALD), the verb *have* has 33 different meanings. The primary meaning of the verb is 'own, hold or possess something'; for example, *He had a new car and a boat* or *Have you got a job yet?*

The dictionary entry of the verb *make* shows 19 different meanings. The first one is 'create or prepare something by combining materials or putting parts together', e.g. *To make a table/dress/cake.* It can also imply 'write, create or prepare something', for example, *These regulations were made to protect children* (Hornby, 2000).

Regarding the verb *take,* the dictionary entry contains 42 different meanings. The first meaning is 'to carry or move something from one place to another', as in *I forgot to take my bag with me when I got off the bus;* 'go with somebody from one place to another', for example, *I'll take you by car*; and 'make somebody/something go from one level, situation, etc. to another', such as *Her energy and talent took her to the top of her profession* (Hornby, 2000).

Taking into account the information provided by Hornby (2000), those collocations in which the verbs under study retain their primary meaning were not

considered in the present study. In other words, only those combinations that were not fully predictable from their component words were selected.

The final step was to check the remaining verb-noun combinations into two dictionaries: *The BBI Dictionary of English Word Combinations* (Benson et al., 1986) and the *Online Oxford Collocation Dictionary of English* (http://oxforddictionary.so8848.com/). If the combination was not included in any of those dictionaries, it was not considered a valid collocation for the study (similar procedures of verification were carried out by Nesselhauf, 2005 and Laufer & Waldman, 2011).

Regarding the BNC corpus, all the occurrences extracted were revised manually, as the web browser shows the main nouns that collocate with the verbs but it does not take into account the words in between. For that reason, some collocates are not valid because *have* acts as an auxiliary verb. For example, *part* appears as one frequent collocate of *have*, but it is not valid for this study since all the occurrences show that *have* acts as an auxiliary verb while the main verb is *take* (e.g. *have taken part*). Once the verb-noun combinations were revised manually, I repeated the process carried out with the SULEC by removing all the expressions in which the verbs retained their primary meaning and those which were not included in the dictionaries mentioned above.

4. Results

4.1. Corpora results

The first research question of the study concerns the main verb-noun collocations of three high-frequency verbs (*take*, *make* and *have*) produced by Spanish learners of English (at different proficiency levels) and native speakers. This section will account for the main results obtained with the three verbs in each of the corpora. After that, results will be analyzed in order to answer the remaining research questions.

4.1.1. SULEC

The structure of this section is as follows. First of all, the results obtained with each individual verb will be presented, including the most frequent collocations used by NS and NNS. Then, the results in the learner corpus will be divided according to the proficiency level of the learners. Finally, the miscollocations found in the learner corpus will be shown.

4.1.1.1. Results with the verb *take*

To begin with, the learner corpus provided 136 occurrences of 33 different types of verb-noun combinations with the verb *take.* Following the criteria described in Section 3.3 above, miscollocations and combinations not relevant to the study were removed from the data, after which the total number of instances extracted became 106 occurrences of 22 different types of collocations[2] (see Table 5 below).

***Take* + noun combinations**	**SULEC**	
	Raw figures	**%**
Miscollocations	9	27.27%
Noncollocations[3]	2	6.06%
Collocations	22	66.67%
TOTAL	**33**	

Table 5: Distribution of *take*+noun combinations (SULEC).

[2] A complete list of the combinations produced can be seen in Appendix I.
[3] Combinations in which the verb retains its primary meaning or not included in the dictionaries of collocations.

Table 6 presents the results of the 22 different collocations with *take* arranged in decreasing frequency and their distribution in the learner corpus. The frequencies in both corpora have been normalized to 100,000 words.

VERB	NOUN[4]	FREQUENCY	
		Raw figures	**Normalized frequency**
Take	*Place*	16	3.56
Take	*Care*	14	3.11
Take	*Decisions*	12	2.67
Take	*Decision*	12	2.67
Take	*Measures*	11	2.44
Take	*Drugs*	8	1.78
Take	*Time*	5	1.11
Take	*Risk*	4	0.89
Take	*Part*	3	0.67
Take	*Degree*	3	0.67
Take	*Control*	2	0.44
Take	*Steps*	2	0.44
Take	*Action*	2	0.44
Take	*Plane*	2	0.44
Take	*Hours*	2	0.44
Take	*Option*	2	0.44
Take	*Exam*	1	0.22
Take	*Look*	1	0.22
Take	*Advice*	1	0.22
Take	*Advantage*	1	0.22
Take	*Bus*	1	0.22
Take	*Exercise*	1	0.22

Table 6: *Take*+noun collocations (SULEC). Normalized frequency per 100,000 words.

As we can see, the most frequent collocations with the verb *take* found in the written discourse of NNS are *take place* 'to happen or occur' (with a frequency of 3.56 per 100,000 words), *take care* 'pay attention; be heedful' (3.11), *take decisions* 'decide' (2.67) and *take (a) decision* (2.67).

[4] Plural form and singular form of the same noun are considered as two different collocations since the meaning of the expression may vary depending on the form of some nouns.

In contrast, we can find some collocations that only appear once in the whole corpus with a normalized frequency (henceforward NF) of 0.22. These single-occurrence items are *take advantage (of)* 'make the most of', *take advice* 'receive guidance', *take (a) bus*, *take (an) exam*, *take exercise* and *take a look* 'look at something'.

The results found were also classified into two proficiency levels, according to the level of competence of the learners. As mentioned above, the learner corpus provided 136 occurrences of 33 different types of combinations with the verb *take*. Intermediate students produced 88 occurrences of 27 different combinations, whereas advanced students produced 48 tokens of 20 types of combinations[5]. Table 7 below presents a summary of the distribution of the combinations produced by learners at both levels, together with the NF per 100,000 words.

***Take* + noun combinations**	**SULEC - Intermediate**		**SULEC - Advanced**	
	Raw figures	**NF**	**Raw figures**	**NF**
Miscollocations	18	7.50	4	1.90
Noncollocations	5	2.08	3	1.43
Collocations	65	27.08	41	19.52
TOTAL	**88**		**48**	

Table 7: Distribution of *take*+noun combinations at different proficiency levels.

Intermediate students outperform advanced students in the production of collocations (27.08 vs. 19.54 respectively). However, advanced students have a lower frequency of errors. Whereas the frequency of miscollocations in intermediate students is 7.5 per 100,000 words, the NF of advanced students is just 1.9.

As shown above, both intermediate and advanced students generate wrong combinations with this verb, such as *take (a) coffee*, *take rules*, *take* (*a*) *lecture*, *take air*, *take* (*an*) *illness*, *take* (*a*) *drink*, *take profits*, *take fun* and *take* (*a*) *change*. Table 8 presents the number of different miscollocations produced and their occurrences.

[5] A complete list of the combinations produced at both proficiency levels can be seen in Appendix I.

	Take
Wrong combinations	9
Occurrences	22

Table 8: Miscollocations with the verb *take*.

Some of these wrong combinations are shown in context in examples (1) to (3) below.

(1) *In a café, the people go for **take fun** and for to relax, too and if this people are used to smoking in this public local when they relax, the question is where can they smoke the cigarettes* (SULEC, Intermediate).[6]

(2) *Smoking is very unhealthy, it can cause cancer or other sickness, if you don't mind to **take** one of this **illness** think in the rest of the people, probably they don't want to be sick neither to perjudicate their lungs* (SULEC, Intermediate).

(3) *The first purpose of the university is to form, to educate a thinking person that will be able to **take profits** of the knowledge that had obtained during the studies* (SULEC, Intermediate).

4.1.1.2. Results with the verb *make*

The second verb under study is *make*. The learner corpus included 128 occurrences of 44 different types of verb-noun combinations with the verb *make*. Nevertheless, a high number of combinations with this verb were considered as miscollocations. Once the data were analyzed, only 23 of the 44 combinations were relevant collocations, according to the criteria mentioned in Section 3.3 above. These are presented in Table 9 below.

[6]Examples extracted from the learner corpus are shown as in the original text, errors included.

***Make* + noun combinations**	**SULEC**	
	Raw figures	**%**
Miscollocations	17	38.64%
Noncollocations	4	9.09%
Collocations	23	52.27%
TOTAL	**44**	

Table 9: Distribution of *make*+noun combinations (SULEC).

The full list of collocations with the verb *make* arranged in decreasing frequency is shown in Table 10 below. As can be seen, there are 23 different types of collocations made up of *make* and a noun.

VERB	**NOUN**	**FREQUENCY**	
		Raw figures	**NF**
Make	*Effort*	10	2.22
Make	*Money*	10	2.22
Make	*Difference*	6	1.33
Make	*Mistake*	6	1.33
Make	*Changes*	4	0.89
Make	*Love*	3	0.67
Make	*Use*	3	0.67
Make	*Decisions*	3	0.67
Make	*Decision*	3	0.67
Make	*Sense*	2	0.44
Make	*Comparison*	2	0.44
Make	*Joke*	2	0.44
Make	*Attempt*	2	0.44
Make	*Distinction*	2	0.44
Make	*Noise*	1	0.22
Make	*Assertion*	1	0.22
Make	*Selection*	1	0.22
Make	*Proposal*	1	0.22
Make	*Discovery*	1	0.22
Make	*Complaint*	1	0.22
Make	*Request*	1	0.22
Make	*Progress*	1	0.22
Make	*Statement*	1	0.22

Table 10: *Make*+noun collocations (SULEC). Normalized frequency per 100,000 words.

As can be seen in Table 10 above, the most frequent collocations with the verb *make* produced by NNS are *make (an) effort* 'try hard' and *make money* 'make a profit', both with a frequency of 2.22 per 100,000 words. Following the former we can find the collocations *make a difference* 'have a significant impact' and *make a mistake* 'commit an error' both with a NF of 1.33.

As was the case with the verb *take*, there are also many single-occurrence items with a frequency of 0.22 per 100,000 words. The following collocations are only produced in one occasion in the whole written corpus: *make (an) assertion*, *make (a) complaint*, *make (a) discovery*, *make (a) noise*, *make progress* 'advance or further something', *make (a) proposal*, *make (a) request* 'ask for something', *make (a) selection* and *make (a) statement*.

From the 128 occurrences of 44 types of verb-noun combinations provided by the learner corpus, intermediate students produced 92 occurrences of 38 types of combinations, whereas advanced students produced 36 occurrences of 20 different combinations. Table 11 shows the distribution of the whole number of verb-noun combinations produced at both proficiency levels together with the NF per 100,000 words.

***Make* + noun combinations**	**SULEC - Intermediate**		**SULEC - Advanced**	
	Raw figures	**NF**	**Raw figures**	**NF**
Miscollocations	27	11.25	7	3.33
Noncollocations	21	8.75	6	2.86
Collocations	44	18.33	23	10.95
TOTAL	**92**		**36**	

Table 11: Distribution of *make*+noun combinations at different proficiency levels.

As with the verb *take*, intermediate students outperform the advanced ones in the production of collocations. Whereas intermediate students show a frequency of 18.33 per 100,000 words, advanced students produce 10.95. Regarding miscollocations, intermediate students also generate more wrong combinations than advanced students (11.22 vs. 3.33 respectively).

The number of wrong combinations produced with this verb is higher than in the previous case, as we can see in Table 12 below which presents the results of the miscollocations produced by learners with the verb *make.*

	Make
Wrong combinations	17
Occurrences	34

Table 12: Miscollocations with the verb *make*.

The seventeen types of wrong combinations extracted from the learner corpus are the following: *make* (*an*) *exam*, *make* (*a*) *question*, *make* (*a*) *demonstration*, *make exercise*, *make pressure*, *make emphasis*, *make diet*, *make* (*an*) *explanation*, *make part*, *make* (*an*) *investigation*, *make fire*, *make business*, *make pain*, *make* (*an*) *opinion*, *make harm* and *make damage*. Some of these can be seen in context in examples (4) to (6) below.

(4) *We should learn to stop kidding about these kind of topics and beging accepting them like part of our society, which we* ***make part*** (SULEC, Intermediate).

(5) *First of all, all the people must have the same rights. And to be married is one of them. Also, it is an action wich doesn't* ***make*** *any* ***pain*** *to nobody* (SULEC, Intermediate).

(6) *They base their opinions in the fact that money is proved to be necessary to achieve happiness and they* ***make*** *special* ***emphasis*** *in the fact that money is needed to achieve happiness but that money is not happiness* (SULEC, Advanced).

4.1.1.3. Results with the verb *have*

Finally, the learner corpus provided 993 occurrences of 86 different types of combinations made up of the verb *have* and a noun. After analyzing the data, a number

of combinations were removed from the list of collocations since they were considered as miscollocations or noncollocations (see Table 13 below).

***Have* + noun combinations**	**SULEC**	
	Raw figures	**%**
Miscollocations	11	12.79%
Noncollocations	10	11.63%
Collocations	65	75.58%
TOTAL	**86**	

Table 13: Distribution of *have*+noun combinations (SULEC).

The results of the collocations formed by the verb *have* and a noun are shown in Table 14. The learner corpus contains 854 occurrences of 65 different collocations made up of the verb *have* and a noun, a much higher number than in the previous cases.

VERB	**NOUN**	**FREQUENCY**	
		Raw figures	**NF**
Have	*Rights*	167	37.11
Have	*Right*	115	25.56
Have	*Problems*	63	14.00
Have	*Children*	59	13.11
Have	*Problem*	34	7.56
Have	*Opportunity*	25	5.56
Have	*Opinion*	25	5.56
Have	*Freedom*	20	4.44
Have	*Knowledge*	18	4.00
Have	*Opportunities*	17	3.78
Have	*Baby*	15	3.33
Have	*Illness*	14	3.11
Have	*Advantages*	14	3.11
Have	*Experience*	14	3.11
Have	*Importance*	14	3.11
Have	*Respect*	13	2.89
Have	*Cancer*	12	2.67
Have	*Consequences*	11	2.44
Have	*Habit*	10	2.22
Have	*Coffee*	10	2.22
Have	*Reasons*	9	2.00
Have	*Idea*	9	2.00
Have	*Disadvantages*	9	2.00
Have	*Value*	8	1.78

Have	*Time*	8	1.78
Have	*Effect*	8	1.78
Have	*Chance*	8	1.78
Have	*Difficulties*	8	1.78
Have	*Feelings*	8	1.78
Have	*Disease*	7	1.56
Have	*Ideas*	7	1.56
Have	*Option*	7	1.56
Have	*Question*	6	1.33
Have	*Role*	6	1.33
Have	*Contact*	5	1.11
Have	*Responsibilities*	5	1.11
Have	*Doubt*	4	0.89
Have	*Fun*	4	0.89
Have	*Accident*	4	0.89
Have	*Addiction*	4	0.89
Have	*Dinner*	4	0.89
Have	*Lunch*	3	0.67
Have	*Influence*	3	0.67
Have	*Choice*	3	0.67
Have	*Reason*	3	0.67
Have	*Prejudices*	3	0.67
Have	*Sex*	3	0.67
Have	*Probabilities*	3	0.67
Have	*Alternative*	3	0.67
Have	*Impression*	2	0.44
Have	*Impact*	2	0.44
Have	*Hope*	2	0.44
Have	*Access*	2	0.44
Have	*Break*	2	0.44
Have	*Consideration*	2	0.44
Have	*Validity*	1	0.22
Have	*Status*	1	0.22
Have	*Beer*	1	0.22
Have	*Chat*	1	0.22
Have	*Belief*	1	0.22
Have	*Dream*	1	0.22
Have	*Trouble*	1	0.22
Have	*Breakfast*	1	0.22
Have	*Relevance*	1	0.22
Have	*Allergy*	1	0.22

Table 14: *Have*+noun collocations (SULEC). Normalized frequency per 100,000 words.

As Table 14 shows, there are two collocations that stand out given their high frequency. *Have rights* is the most common collocation found in the corpus with a NF of 37.11, followed by *have (a) right* with a frequency of 25.56 per 100,000 words. In third and fourth positions we can find *have problems* (with a NF of 14.00) and *have children* (13.11).

Similarly to the previous cases, there is a remarkable number of collocations that appears only once, with a NF of 0.22. The single occurrence items found with the verb *have* are the following: *have (an) allergy*, *have (a) beer*, *have (a) belief*, *have breakfast* 'have morning meal', *have (a) chat*, *have (a) dream*, *have relevance*, *have (a) status*, *have trouble* 'experience difficulty doing something' and *have validity*.

Regarding the level of competence of the students, the learner corpus provided 993 occurrences of 86 different types of combinations with the verb *have* from which intermediate students produced 697 occurrences of 76 different combinations, whereas advanced students produced 296 occurrences of 61 types of combinations. Table 15 presents a summary of the verb-noun combinations produced with the verb *have* by both intermediate and advanced students.

***Have* + noun combinations**	**SULEC – Intermediate**		**SULEC – Advanced**	
	Raw figures	**NF**	**Raw figures**	**NF**
Miscollocations	20	8.33	3	1.43
Noncollocations	71	29.58	45	21.43
Collocations	606	252.5	248	118.1
TOTAL	**697**		**296**	

Table 15: Distribution of *have*+noun combinations at different proficiency levels.

As Table 15 shows, intermediate students surpass advanced students in the production of collocations once again (252.5 vs. 118.1 respectively). Similar results were obtained for the frequency of miscollocations. While intermediate students generate wrong combinations with a frequency of 8.33 per 100,000 words, advanced students showed a NF of 1.43. In general, the production of verb-noun combinations of the intermediate group is higher. However, the errors they make are also more numerous than those of the advanced group.

Regarding miscollocations, Table 16 shows the total number of wrong combinations produced by learners and their occurrences. The wrong combinations produced with this verb are the following: *have sense, have years, have* (*a*) *punishment, have* (*a*) *war, have guilt, have health, have drugs, have care, have blame, have control* and *have* (*a*) *risk*.

	Have
Wrong combinations	11
Occurrences	23

Table 16: Miscollocations with the verb *have*.

Some of these combinations can be seen in context in examples (7) and (9) below.

(7) *The Government* ***have*** *the* ***guilt*** *because it earns a lot of money and the companies that go on selling tobacco instead of it is a bad thing* (SULEC, Intermediate).

(8) *It's like a situation where the yonkis reclamate theyr right to* ***have drugs*** *in a public place, it sounds mad, but they are really covered by a marginal situation and they are not accepted* (SULEC, Advanced).

(9) *I accept that people smoke, but I don't like tobacco and I want* ***have health*** (SULEC, Intermediate).

To sum up, Table 17 presents a comparison on the NF of the three verbs under study in the learner corpus. As we can see, the NF of the collocations with the verb *have* is much higher than in the other cases, which means that learners produce more verb-noun collocations with this verb.

	Types	**Tokens**	**NF**
Take	22	106	23.56
Make	23	67	14.89
Have	65	854	189.78
TOTAL	**110**	**1027**	

Table 17: Normalized frequency of collocations with the three verbs under study (SULEC).

4.1.2. BNC

This section will present the main results obtained from the corpus of native speakers. Given its extension, the BNC contains a higher number of words that collocate with the verbs under study. As the large amount of data included in this corpus makes it almost impossible to extract and check all the available collocations with the verbs under study, only the 20 most frequent verb-noun combinations were extracted.

Table 18 below shows the results of the 20 most frequent *take*+noun collocations in the native corpus. Frequencies have been normalized to 100,000 words in order to obtain comparable data with the learner corpus.

VERB	**NOUN**	**FREQUENCY**	
		Raw figures	**NF**
Take	*Place*	2369	15.45
Take	*Account*	435	2.84
Take	*View*	387	2.52
Take	*Form*	385	2.51
Take	*Part*	361	2.35
Take	*Time*	285	1.86
Take	*Action*	250	1.63
Take	*Steps*	243	1.58
Take	*Advantage*	241	1.57
Take	*Care*	228	1.49
Take	*Forms*	149	0.97
Take	*Responsibility*	147	0.96
Take	*Effect*	136	0.89
Take	*Example*	110	0.72
Take	*Years*	98	0.64
Take	*Interest*	83	0.54
Take	*Step*	69	0.45
Take	*Possession*	68	0.44
Take	*Precedence*	65	0.42
Take	*Decisions*	58	0.38

Table 18: *Take*+noun collocations (BNC). Normalized frequency per 100,000 words.

As we can see, the commonest combination among native speakers is *take place* standing out from other collocations with a NF of 15.45. In the second place we find the collocation *take account (of)* 'take into consideration' but its NF (2.84) is far from that of the first one. It is followed by *take (a) view* and *take (the) form (of)*, whose frequency per 100,000 words is quite similar (2.52 vs. 2.51 respectively). The last combination extracted is *take decisions* in the 20th position, with a much lower NF (0.38).

Similarly, Table 19 below shows the 20 most frequent nouns that collocate with the verb *make* in the native corpus.

VERB	**NOUN**	**FREQUENCY**	
		Raw figures	**NF**
Make	*Use*	726	4.74
Make	*Sense*	610	3.98
Make	*Decisions*	288	1.88
Make	*Contribution*	259	1.69
Make	*Difference*	250	1.63
Make	*Decision*	227	1.48
Make	*Point*	226	1.47
Make	*Provision*	161	1.05
Make	*Distinction*	149	0.97
Make	*Representations*	111	0.72
Make	*Attempt*	109	0.71
Make	*Reference*	100	0.65
Make	*Statement*	96	0.63
Make	*Progress*	94	0.61
Make	*Changes*	92	0.60
Make	*Choice*	91	0.59
Make	*Claims*	91	0.59
Make	*Application*	86	0.56
Make	*Demands*	81	0.53
Make	*Statements*	80	0.52

Table 19: *Make*+noun collocations (BNC). Normalized frequency per 100,000 words.

As can be seen, the first two collocations stand out with a higher NF than the others. In the first position we find *make use (of)* 'utilize' with a NF of 4.74, whereas *make sense* 'be logical' comes in the second place with a NF of 3.98. *Make decisions* appears in the third position right followed by *make (a) contribution* with normalized frequencies of 1.88 and 1.69 respectively. At the opposite side the collocation *make statements* appears in the last place, with a NF of 0.52.

Finally, Table 20 below presents the results of the 20 most frequent nouns that collocate with *have* in the native corpus.

VERB	**NOUN**	**FREQUENCY**	
		Raw figures	**NF**
Have	*Effect*	1,588	10.36
Have	*Right*	678	4.42
Have	*Power*	673	4.39
Have	*Access*	385	2.51
Have	*Impact*	369	2.41
Have	*Interest*	366	2.39
Have	*Role*	341	2.22
Have	*Influence*	337	2.20
Have	*Implications*	332	2.17
Have	*Experience*	291	1.90
Have	*Regard*	285	1.86
Have	*Effects*	267	1.74
Have	*Knowledge*	260	1.70
Have	*Opportunity*	259	1.69
Have	*Difficulty*	258	1.68
Have	*Control*	243	1.58
Have	*Meaning*	234	1.53
Have	*Reason*	234	1.53
Have	*Authority*	224	1.46
Have	*Children*	215	1.40

Table 20: *Have*+noun collocations (BNC). Normalized frequency per 100,000 words.

Have (an) effect 'to make an impact' is the commonest collocation with this verb in the written discourse of native speakers with a NF of 10.36. Right following it, we

find *have (a) right* 'be entitled', whose NF is 4.42, and *have power* 'have control or influence', with a frequency of 4.39 per 100,000 words. On the opposite side, *have children* appears in the last position with a NF of 1.40.

4.2. Discussion of findings

4.2.1. Differences in collocational production

In Section 4.1 above we gave answer to the first research question addressed in the present piece of research, namely what are the main verb-noun collocations produced by Spanish learners and native speakers with the three verbs under study (*take, make* and *have)*.

The object of this section is to answer the second research question, which concerns the differences in the collocational production of NNS and NS, for which the data obtained from both corpora are compared. Although there are striking differences in the production of collocations in the two corpora, results also show some similarities. First of all, comparisons of the results of individual verbs will be shown and then we will provide a summary of the general differences between the findings in both corpora.

Regarding the verb *take*, comparing Tables 6 and 18 above, we can observe that from the 21 collocations found in the NNS corpus, only 8 appear in the list of the corpus of NS. These include the following: *take action*, *take advantage (of*), *take care*, *take part*, *take place*, *take (a) step* and *take time*. Although the data obtained from the SULEC are scarce and should probably be complemented by further data in future research, at first sight it seems that Spanish learners make more frequent use of certain combinations which are not so common among the NS, such as *take advice*, *take exercise*, *take hours*, *take measures*, *take (a) risk*, etc. For example, in the BNC *take hours* has a NF of 0.17 (as opposed to a frequency of 0.44 in the SULEC), whereas *take exercise* displays a frequency of 0.05 per 100,000 words (as opposed to 0.22 in the SULEC). In contrast, learners do not use collocations that have a high frequency among NS of the language, such as, e.g. *take account (of), take effect*, *take (the) form (of), take possession* and *take (a) view*.

Nevertheless, it is interesting to note that the most frequent collocation with the verb *take* is the same in both corpora, namely *take place*. NNS produced this collocation 16 times (a frequency of 3.56 per 100,000 words) whereas the corpus of NS provided 2,369 tokens (a frequency of 15.45 per 100,000 words). As we can see, the production of this collocation is quite higher in the case of NS in comparison to NNS, which means that learners seem to underuse this collocation as compared to NS.

In fact, learners tend to underuse most collocations in relation to NS. If we compare the collocations which appear in the two corpora, we can see that their frequency is much lower in the case of learners than in the case of NS, as shown in Table 21 below.

	SULEC		**BNC**	
	NF	**Position**	**NF**	**Position**
Take action	0.44	13th	1.63	7th
Take advantage	0.22	19th	1.57	9th
Take part	0.67	9th	2.35	5th
Take place	3.56	1st	15.45	1st
Take steps	0.44	12th	1.58	8th
Take time	1.11	7th	1.86	6th

Table 21: Comparison of the use of *take*+noun collocations between NNS and NS.

Some of the differences in the production of certain types of collocations are quite remarkable, as in the case of *take place*, *take part*, *take steps*, *take action* and *take advantage*, since the frequencies are much higher in the corpus of NS. However, in the case of *take time* the difference is not as striking as in the previous cases.

At the other end of the spectrum, we find collocations in the learner corpus which are much more frequent than in the native corpus. The second and third most frequent collocations in the SULEC (*take care* and *take decisions*) have a higher frequency than in the BNC. *Take care* has a frequency of 3.11 in the SULEC (14 tokens) while a frequency of 1.49 in the BNC (228 tokens). In this case, the NF is higher in the learner corpus, which means that learners tend to overuse this kind of collocation. The same happens with the collocation *take decisions*, whose frequency in the learner corpus is 2.67 (12 occurrences) per 100,000 words whereas in the BNC it appears in the 20th position with a NF of 0.38 (58 occurrences).

One of the reasons that may account for the overuse of the collocation *take decisions* is that there is a word-for-word translation in Spanish. The Spanish expression *tomar decisiones* is literally translated into English as *take decisions*. As a consequence, it is a relatively easy collocation for Spanish learners. Nevertheless, other collocations produced, such as *take part*, also have a literal translation in Spanish (*tomar parte*) but they are not overused by learners. Therefore, further research would be necessary in order to determine the factors that account for the overuse of certain collocations.

Another reason could be that the topics of the SULEC essays elicit that kind of collocation. Most occurrences of this collocation in the learner corpus are related to the topic of the smoking ban in public places. Some learners use the expression many times to claim, on the one hand, that the government has decided to impose an anti-smoking law without asking people's opinion, as in example (10) below.

(10) *Nowadays, smoke it is an activity that it is being very discussing by government and authorities. They are* ***taking decisions*** *without ask people* (SULEC, Advanced).

On the other hand, the expression is used to talk about the rights of non-smokers to be in a public place without being exposed to tobacco smoke. Example (11) below is another instance extracted from the corpus.

(11) *Smoking is not something that you do just for and to yourself but its smoke expands and spreads all over a room in which non-smokers will be passively involved. It is precisely because of this reason, that smokers can not* ***take decisions*** *for non-smokers* (SULEC, Advanced).

In the case of *take care* it is not so clear why learners tend to overuse this collocation in comparison to NS. The expression is more idiomatic than *take decisions* so that there is no possibility to establish a direct translation from Spanish. However, if we have a look at the contexts in which the collocation is found, we can observe that most of them refer to one of the topics of the essays included in the SULEC, namely the learners' opinion about marriage of persons of the same sex. Most learners use the expression *take care* to convey the idea that two people of the same sex can look after children in the same way as a man and a woman do, as illustrated in example (12) below.

(12) *On the one hand, from my point of view, they can* ***take care*** *of children as well as other people, they can love them, buy them, buy them the things that they need, feed them...*(SULEC, Intermediate).

Therefore, the most plausible explanation is that the collocation is used by learners as a consequence of the topic of the essay, as in the case of *take decisions*. If the topics chosen were different, maybe these collocations would not be as common as they are now in the corpus.

Regarding the verb *make*, the lists of collocations extracted from both corpora and illustrated in Tables 10 and 19 above allow us to establish some similarities and differences in the production of NNS and NS. From the 23 different types of collocations found in the SULEC, only 10 of them are included among the 20 most frequent collocations in the BNC, namely *make (an) attempt*, *make changes*, *make (a) decision*, *make decisions*, *make (a) difference*, *make (a) distinction*, *make progress*, *make sense*, *make(a) statement* and *make u*se *(of)*. The remaining collocations used by learners are not as frequent in the written discourse of NS (e.g. *make (an) effort*, *make (an) exam*, *make (a) joke*, *make love*, *make (a) mistake* and *make money*). As an example, *make (a) joke* has a NF of 0.06 in the BNC, whereas its frequency in the SULEC is 0.44. At the opposite side, there are many collocations commonly used by NS (e.g. *make claims*, *make (a) contribution*, *make (a) point*, *make provision (for)*, *make representations*) that are not found in the learner corpus.

From the 10 *make*+noun collocations found in both lists, 9 of them are underused by learners in comparison to data extracted from the NS corpus. Table 22 presents the frequency of collocations underused by learners and their position according to the number of occurrences of the combination.

	SULEC		BNC	
	NF	**Position**	**NF**	**Position**
Make (an) attempt	0.44	14th	0.71	11th
Make (a) decision	0.67	10th	1.48	6th
Make decisions	0.67	9th	1.88	3rd
Make (a) difference	1.33	3rd	1.63	5th
Make (a) distinction	0.44	15th	0.97	9th
Make progress	0.22	23rd	0.61	14th
Make sense	0.44	11th	3.98	2nd
Make (a) statement	0.22	24th	0.63	13th
Make use	0.67	8th	4.74	1st

Table 22: Comparison of the use of *make*+noun collocations between NNS and NS.

There are some remarkable differences between the frequency of certain collocations in the native and the learner corpus. *Make use* is the most frequent collocation produced by NS with a frequency of 4.74. However, it appears in the 8th position in the list of collocations of NNS, with a much lower frequency (0.67). The same happens with the collocation *make sense*, very common in the native corpus (3.98) but not in the corpus of learners, where it shows a NF of 0.44.

In contrast to these findings, there is one remaining collocation that both corpora have in common but is most frequently used by learners, namely *make changes*. This collocation is more commonly found in the NNS corpus than in the NS one (0.89 vs. 0.60 respectively). Although the difference in frequencies is slight, a priori it seems that learners tend to overuse that particular combination in contrast to the underuse of the rest of collocations.

In previous cases, collocations which were overused with the verb *take* (i.e. *take care* and *take decision*s) were related to a specific topic of the learner corpus. However, the collocation *make changes* is not restricted to one specific topic, it can be found in essays about different issues. As a consequence, the overuse of this collocation cannot be explained in terms of the topics chosen. However, this structure has a parallel translation in Spanish. The expression *hacer cambios* is literally translated into English as *make changes.* Therefore, as it happened with *take decisions*, it may be a relatively easy collocation to produce by Spanish learners given their similarity to the Spanish equivalent.

Finally, collocations with *have* are the most frequent type of combinations found in the learner corpus in contrast to the verbs *take* and *make* (see Table 17 above). Comparing Tables 14 and 20 above, where the lists of collocations from the learner and native corpus are displayed, we can observe that 11 out of the total 65 collocations found in the learner corpus are included among the 20 most frequent collocations in the BNC: *have access*, *have children*, *have* (*an*) *effect*, *have experience*, *have* (*an*) *impact*, *have influence*, *have knowledge*, *have* (*an*) *opportunity*, *have* (*a*) *reason*, *have* (*a*) *right* and *have* (*a*) *role*. Most collocations learners use are not particularly frequent in the written discourse of NS (e.g. *have (an) idea*, *have (an) illness*, *have (an) opinion*, *have problems*, *have respect*, *have value*, etc.). On the other hand, there are high frequency collocations in the BNC corpus that are not used in the written production of Spanish learners (e.g. *have authority*, *have control*, *have implications (for)*, *have power*, *have regard (to)*, etc.).

Regarding those collocations that are common to both corpora, some of them are used with a higher frequency by NS but some other are more common in the written discourse of NNS. Table 23 below presents the results of the collocations found in both corpora that are underused by learners, while Table 24 shows those collocations that are overused by NNS of English.

	SULEC		**BNC**	
	NF	**Position**	**NF**	**Position**
Have access	0.44	53rd	2.51	4th
Have (an) *effect*	1.78	26th	10.36	1st
Have (*an*) *impact*	0.44	51st	2.41	5th
Have influence	0.67	43rd	2.20	8th
Have (*a*) *reason*	0.67	45th	1.53	18th
Have (*a*) *role*	1.33	34th	2.22	7th

Table 23: Comparison of the use of *have*+noun collocations between NNS and NS. Combinations underused by learners.

	SULEC		BNC	
	NF	**Position**	**NF**	**Position**
Have children	13.11	4th	1.40	20th
Have experience	3.11	14th	1.90	10th
Have knowledge	4.00	9th	1.70	13th
Have (*an*) *opportunity*	5.56	6th	1.69	14th
Have (*a*) *right*	25.56	2nd	4.42	2nd

Table 24: Comparison of the use of *have*+noun collocations between NNS and NS. Combinations overused by learners.

The collocation *have (an) effect* 'cause a result in someone or something' is the most frequent combination in the written production of NS with a frequency of 10.36. It is also produced by learners but its frequency is quite lower (1.78), which indicates that it is a clearly underused combination. In general, all the combinations that can be seen in Table 23 show remarkable differences in terms of frequency between the two corpora.

In contrast, we can find some collocations that are more commonly found in the written production of learners. As it is shown in Table 24 above, there are remarkable differences regarding frequency between five particular collocations in both corpora. It is interesting to note some examples, such as *have (a) right* 'have a just or legal claim on something or on some action'. This combination is the second most frequent collocation found in both corpora. However, its frequency is more than five times higher in the SULEC than in the BNC. It is a striking difference that shows a clear overuse of certain combinations with the verb *have* by learners, which accounts for the global overuse of the structures with this verb.

The use of this particular collocation is undoubtedly influenced by the topic of the essays in the learner corpus. A remarkable percentage of the occurrences of the collocation *have (a) right* correspond to the topic of marriage of people of the same sex, already mentioned above. Most learners use this expression to claim that homosexual people should be able to get married in the same way as a man and a woman do. Example (13) below shows an instance in context from the learner corpus.

(13) *I think that these couples **have the right** to be married and that they also deserves the opportunity to be happy* (SULEC, Intermediate).

In addition, this expression has a word-for-word translation in the learners' mother tongue (*tener el derecho de*), which means that it may be easy for learners to produce this kind of combination.

Overall, the results obtained with the three verbs show that NNS production of collocations is different from that of NS. In the three cases learners underuse certain collocations while they overuse others in comparison to NS's norm. These results are similar to previous studies that point out that learners underuse native-like collocations, such as combinations with the intensifier *highly* (Granger, 1998) but, on the other hand, overuse of some collocations in situations where more specific expressions are required (Shih, 2000).

Some possible explanations that would account for the overuse of certain collocations have been mentioned above. However, it is interesting to note that the total number of collocations overused by learners (*have (a) right*, *have children*, *have (an) opportunity*, *have knowledge* and *have experience*) possess a word-for-word translation in Spanish. Therefore, apart from the influence of the topics given in the corpus, the similarity between L1 and L2 structures may be the main reason why learners tend to use those collocations more frequently. These results are in line with the Contrastive Analysis approach (Lado, 1957), which refers to the idea that, if a structure in the L1 is similar to that in the L2, the consequence is a positive transfer. It means that learners find it easier to produce collocations that have a literal Spanish equivalent.

In contrast, the underuse of certain collocations by learners cannot be explained as a consequence of the similarities and differences between the L1 and L2. There is not a fixed pattern which explains which expressions are underused and which ones are overused. The collocations underused by learners vary from very idiomatic expressions (e.g. *make sense* or *take place*) to collocations that can be translated directly into Spanish, such as *have (a) reason* or *take part*. Hence, the most plausible explanation to account for the underuse of some collocations in comparison to NS is that the topics given did not favor the use of those expressions.

4.2.2. Collocational competence and level of proficiency

The third research question addressed in the study is concerned with the relationship between the proficiency level of learners and their production of collocations. Since the written data in the learner corpus is divided into three levels of competence it was possible to analyze all the data extracted from this perspective. The first idea of the study was to classify all the combinations extracted into those three levels: elementary, intermediate and advanced. However, when looking for examples of the three verbs in the elementary section, I noticed there was little data corresponding to this level. There were no occurrences of the verbs *take* and *make* and just one combination was extracted with the verb *have* (*have a right*). Given the lack of results, the elementary level was not taken into account in the study.

In this section, we will first analyze the differences in the production of collocations with the three verbs under study according to the proficiency level of the learners. Then, we will provide a general overview of the results at both levels in order to establish a relationship between the level of proficiency and the collocational knowledge.

As mentioned in Section 4.1.1.1, the learner corpus provided 136 occurrences of 33 different types of combinations with the verb *take*. Intermediate students produced 88 occurrences of 27 different combinations whereas advanced students produced 48 occurrences of 20 types of combinations. A summary of the distribution of combinations produced by learners at both levels was shown in Table 7 above, repeated here as Table 25.

***Take* + noun combinations**	**SULEC - Intermediate**		**SULEC - Advanced**	
	Raw figures	NF	Raw figures	NF
Miscollocations	18	7.50	4	1.90
Noncollocations	5	2.08	3	1.43
Collocations	65	27.08	41	19.52
TOTAL	**88**		**48**	

Table 25: Distribution of *take*+noun combinations at different proficiency levels.

Figure 5 below shows a comparison in the production of *take*+noun combinations at intermediate and advanced level. As we can see, the NF of combinations among intermediate students is higher with the three possible cases. In other words, intermediate students produce more miscollocations and noncollocations, but also a higher number of collocations than advanced students.

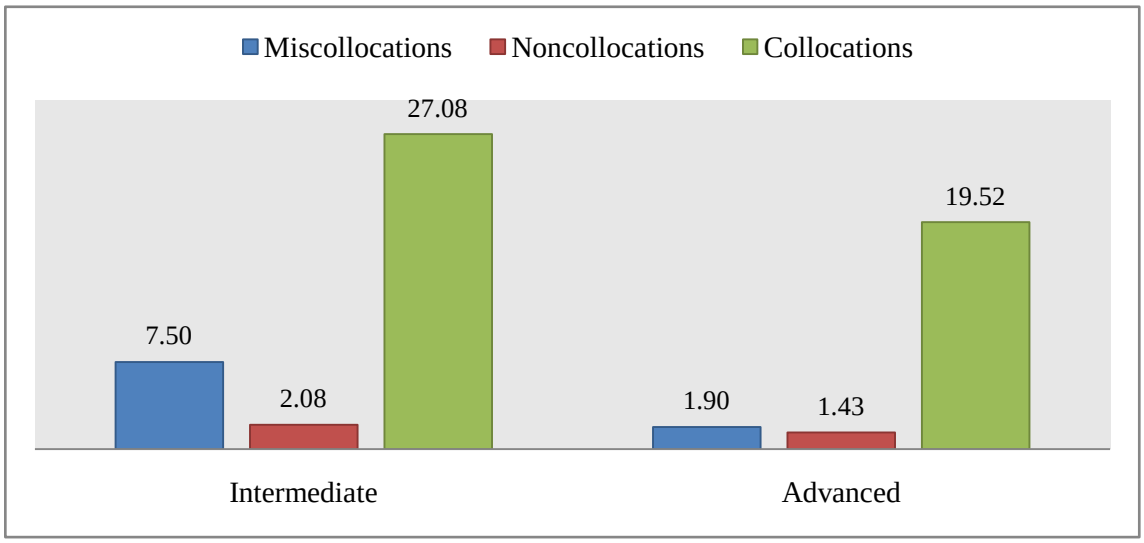

Figure 5: *Take*+noun combinations at different proficiency levels. Normalized frequency per 100,000 words.

As shown in Figure 6 below, from the 88 occurrences of the 27 different types of combinations produced by intermediate students, 20.45% correspond to miscollocations, 5.68% to noncollocations and the remaining 73.86% represents the percentage of collocations.

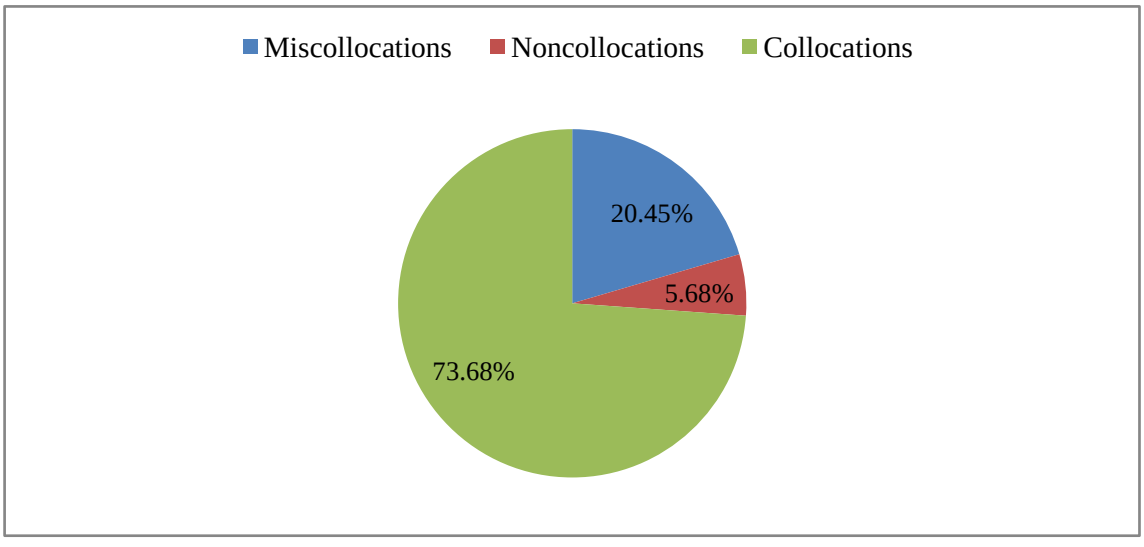

Figure 6: *Take*+noun combinations at intermediate level.

On the other hand, as displayed in Figure 7 below, from the 48 occurrences of 20 different collocations produced by advanced students, 8.33% represent miscollocations, 6.25% noncollocations, whereas the remaining 85.41% correspond to proper collocations.

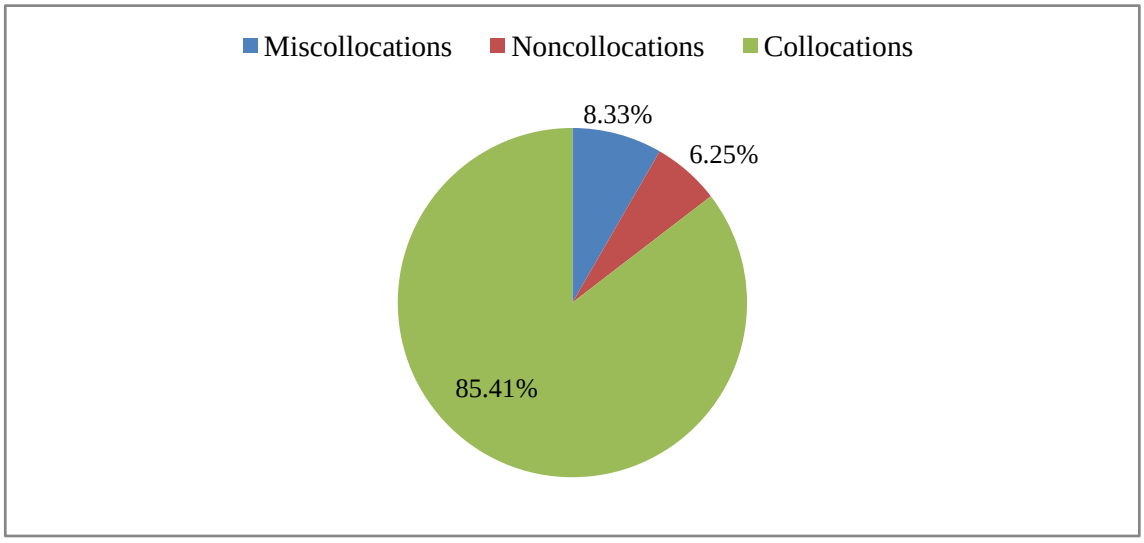

Figure 7: *Take*+noun combinations at advanced level.

With regard to *make*, 128 occurrences of 44 types of verb-noun combinations were found in the corpus. Intermediate students produced 92 occurrences of 38 types of combinations while advanced students produced 36 occurrences of 20 combinations. The distribution of the whole number of verb-noun combinations produced at both proficiency levels was displayed in Table 11above, repeated here as Table 26.

***Make* + noun combinations**	**SULEC - Intermediate**		**SULEC - Advanced**	
	Raw figures	**NF**	**Raw figures**	**NF**
Miscollocations	27	11.25	7	3.33
Noncollocations	21	8.75	6	2.86
Collocations	44	18.33	23	10.95
TOTAL	**92**		**36**	

Table 26: Distribution of *make*+noun combinations at different proficiency levels.

The results for both groups are compared in Figure 8 below. The outcome is similar to that with the verb *take*. In general, intermediate students produce more verb-noun combinations than advanced students. The NF of collocations is higher in the intermediate group than in the advanced one (18.33 vs. 10.95 respectively), as well as the number of miscollocations and noncollocations. The number of wrong combinations produced by learners at both levels is especially remarkable since it is much higher than in the previous case with the verb *take*.

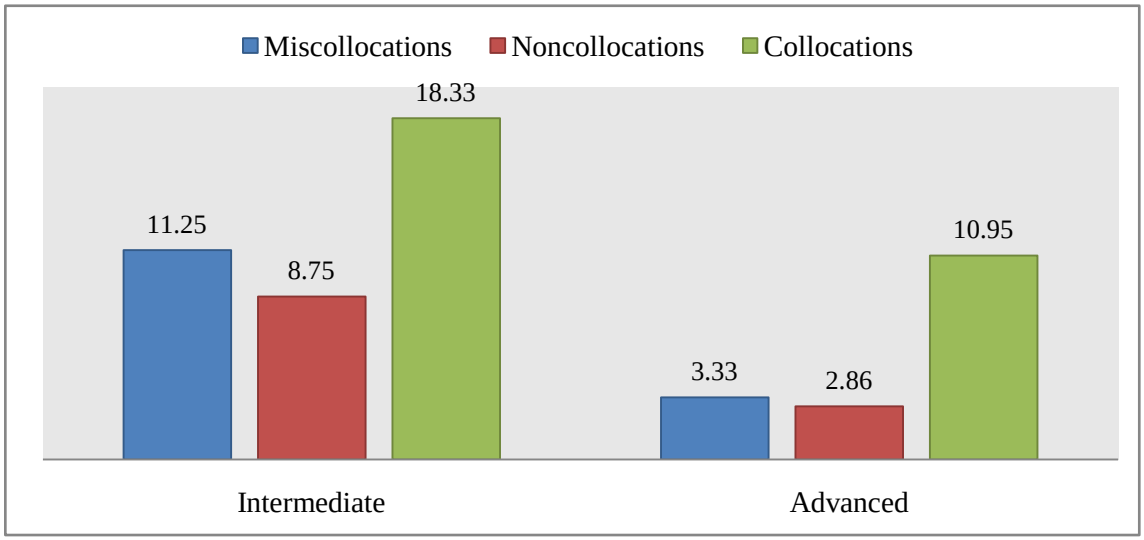

Figure 8: *Make*+noun combinations at different proficiency levels. Normalized frequency per 100,000 words.

As shown in Figure 9 below, the proportion of combinations among intermediate students is the following: 29.34% are miscollocations, 22.83% noncollocations and 47.83% correspond to collocations.

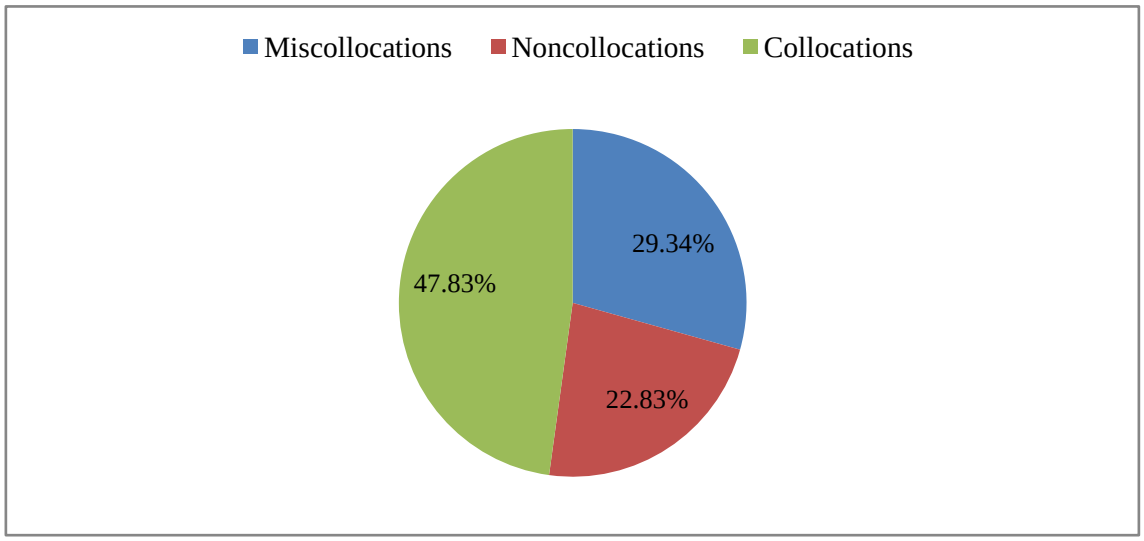

Figure 9: *Make*+noun combinations at intermediate level.

It is interesting to note that less than half of verb-noun combinations produced are collocations while a high percentage include wrong combinations. Given the noticeable percentage of wrong combinations, some illustrative examples of miscollocations are offered in (14) to (16) below.

(14) *Anyway, in my opinion, all people related to university, either professors and pupils, should* ***make pressure*** *in order to get to all these things we need so much; and why not starting with practical courses at the end of the studies?* (SULEC, Intermediate).

(15) *Everybody know that* ***make*** *hard* ***diet****, eat without keeping a medical control and try to be another person is unweathy but the people do it* (SULEC, Intermediate).

(16) *I will try to* ***make*** *a brief but clear* ***explanation*** *about that* (SULEC, Intermediate).

Regarding the advanced group, 19.44% of the verb-noun combinations are miscollocations, 16.67% noncollocations and 63.89% collocations, as can be seen in Figure 10 below.

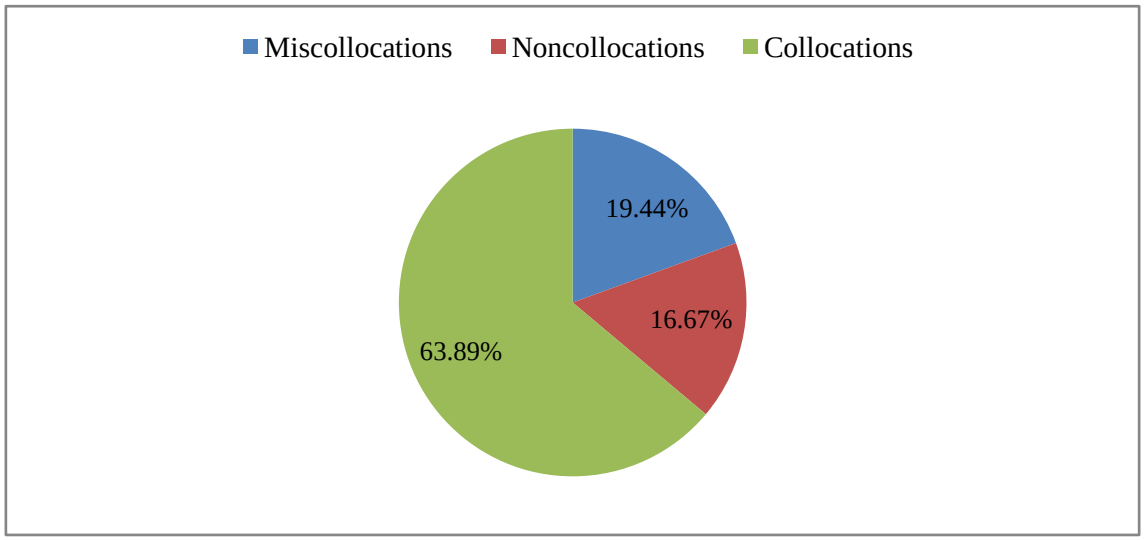

Figure 50: *Make*+noun combinations at advanced level.

The percentage of collocations produced is higher than in the case of intermediate students; however, the percentage of miscollocations at the advanced level

is also representative since almost one out of five combinations produced is wrong. Below are some examples of wrong combinations extracted from the corpus (see (17) and (18)).

(17) *They base their opinions in the fact that money is proved to be necessary to achieve happiness and they **make** special **emphasis** in the fact that money is needed to achieve happiness but that money is not happiness* (SULEC, Advanced).

(18) *I agree about bringing an opportunity to the homosexual persons because it is something normal in our days; they are persons too!!they don't **make harm** to nobody!!* (SULEC, Advanced).

Finally, the learner corpus provided 993 occurrences of 86 different types of combinations with the verb *have*. Intermediate students produced 697 occurrences of 76 different combinations, whereas advanced students produced 296 occurrences of 61 types of combinations. These results were displayed in Table 15 above, repeated here as Table 27.

***Have* + noun combinations**	**SULEC – Intermediate**		**SULEC – Advanced**	
	Raw figures	**NF**	**Raw figures**	**NF**
Miscollocations	20	8.33	3	1.43
Noncollocations	71	29.58	45	21.43
Collocations	606	252.5	248	118.1
TOTAL	**697**		**296**	

Table 27: Distribution of *have*+noun combinations at different proficiency levels.

As in the previous cases with the verbs *take* and *make*, the production of intermediate students is higher for collocations, noncollocations and miscollocations. The NF of collocations produced by the intermediate group is twice the frequency of the advanced group. Another remarkable fact is that there is a great difference between the NF of collocations and miscollocations in both groups. The NF of collocations among intermediate students is 252.5, while that of miscollocations is just 8.33 per 100,000

words. The same happens with the advanced group where the NF of collocations is 118.1, whereas the NF of miscollocations is much lower (1.43).

Figure 11 below shows a comparison in the production of *have*+noun combinations by intermediate and advanced students.

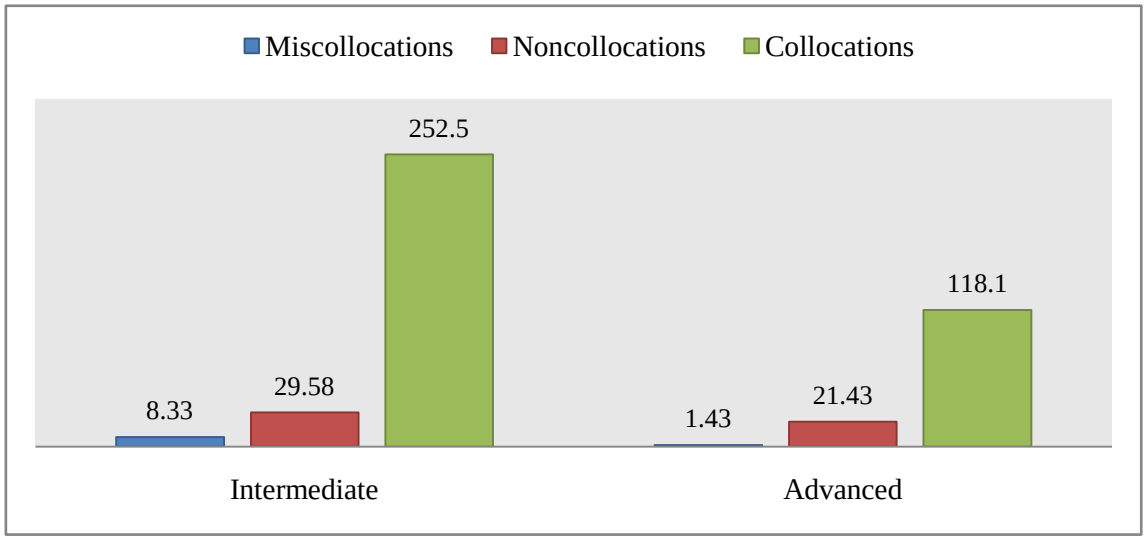

Figure 11: *Have*+noun combinations at different proficiency levels. Normalized frequency per 100,000 words.

Overall, from the 697 occurrences produced by intermediate students, 2.87% are miscollocations, 10.19% noncollocations and the remaining 86.94% are collocations (see Figure 12 below). It is noteworthy that the percentage of miscollocations with this verb is lower than that of the verbs *take* and *make*; as a consequence, the percentage of collocations with *have* is the highest one.

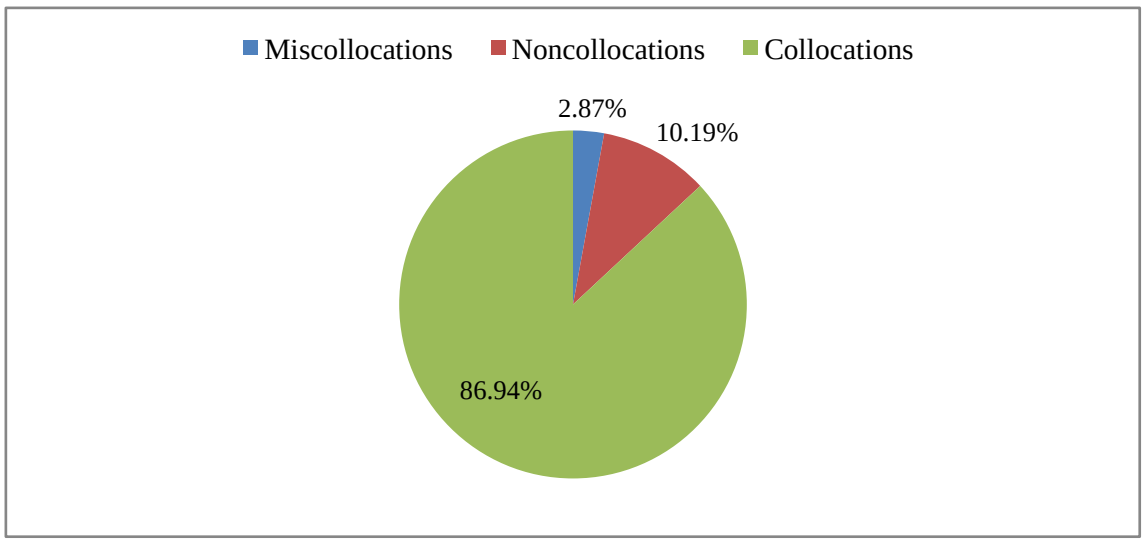

Figure 12: *Have*+noun combinations at intermediate level.

The advanced group produced 296 occurrences, from which 1.01% are miscollocations, 15.20% noncollocations and 83.78% collocations, as represented in Figure 13 below.

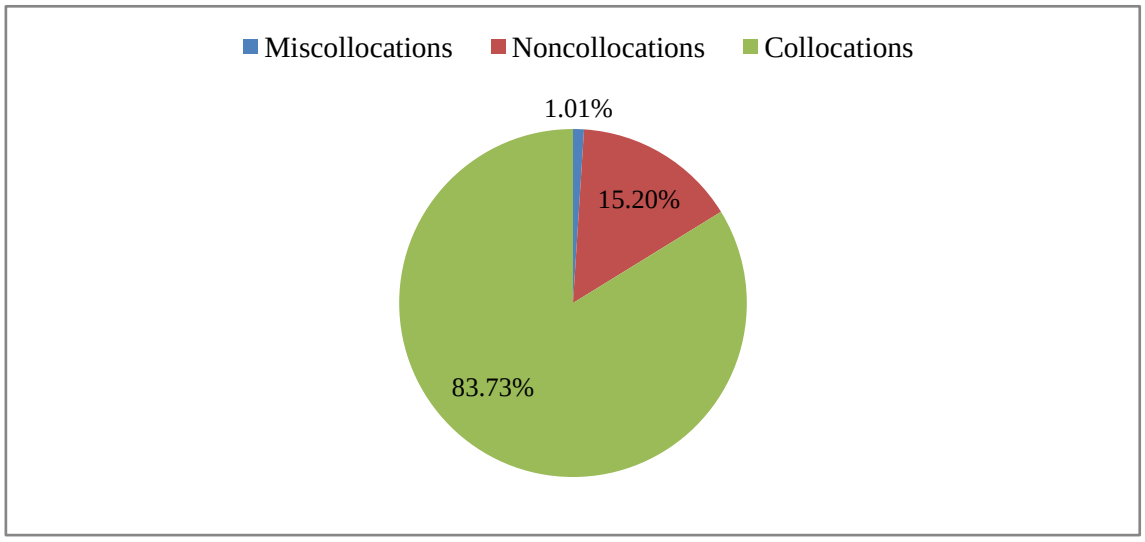

Figure 13: *Have*+noun combinations at advanced level.

In this case, the percentage of collocations is quite higher but not the highest one, given that this group obtained 85.41% of correct collocations with the verb *take*. However, the percentage of miscollocations represents an almost negligible 1.01%.

Overall, intermediate students produce more verb-noun combinations than advanced students as far as collocations, miscollocations and noncollocations are concerned. The NF of collocations is always higher in the case of the intermediate group. However, intermediate learners also produce more wrong combinations than advanced students. Table 28 below presents the general results of both groups regarding the production of miscollocations, noncollocations and collocations with the three verbs.

Verb+ noun combinations	**SULEC - Intermediate**		**SULEC - Advanced**	
	Raw figures	**NF**	**Raw figures**	**NF**
Miscollocations	65	27.08	14	6.67
Noncollocations	97	40.42	54	25.71
Collocations	715	297.92	312	148.57
TOTAL	**877**		**380**	

Table 28: Distribution of verb-noun combinations with the verbs *take*, *make* and *have* at different proficiency levels.

As can be seen, the NF of collocations produced by intermediate students is twice the NF of advanced students. In contrast, the frequency of wrong combinations produced by intermediate students is almost five times the NF of errors made by advanced students. Figure 14 displays a comparison on the NF of verb-noun combinations produced at both proficiency levels.

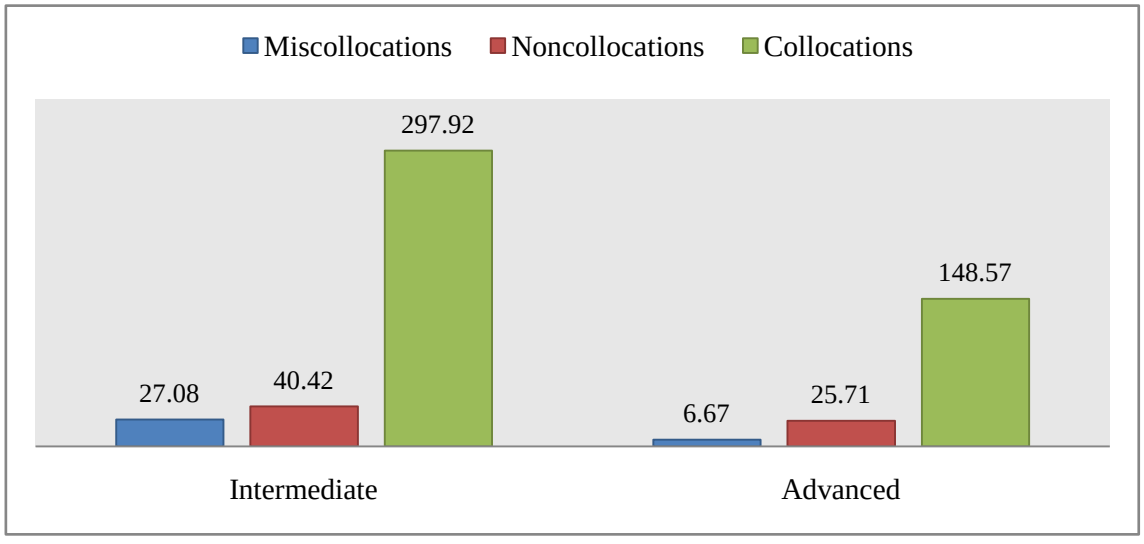

Figure 14: Verb-noun combinations at different proficiency levels. Normalized frequency per 100,000 words.

The intermediate group produced 877 verb-noun combinations with the three verbs under study, out of which 7.41% are miscollocations, 11.06% noncollocations and 81.53% collocations, as shown in Figure 15 below.

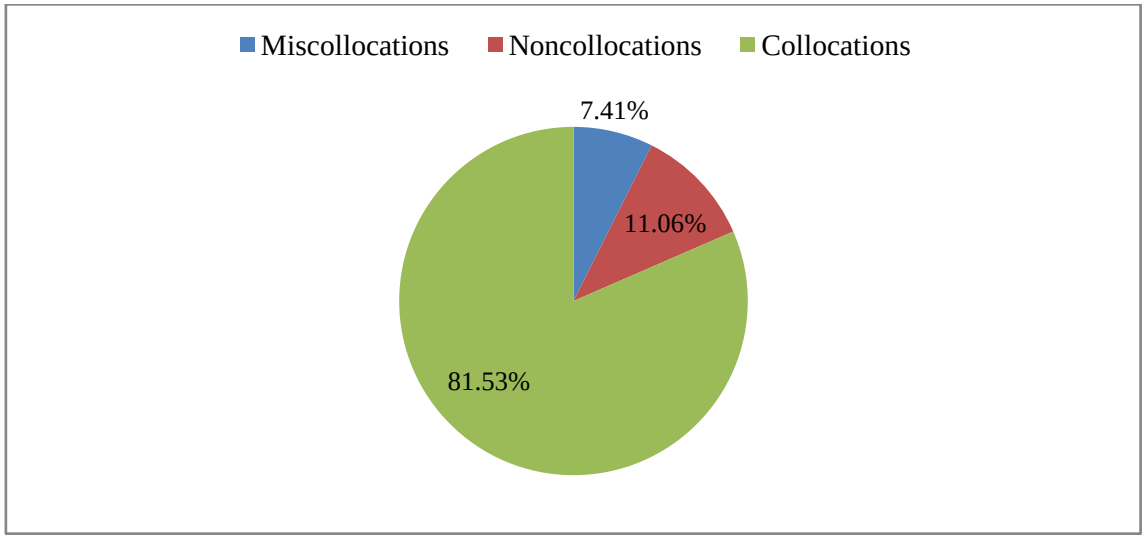

Figure 15: Verb-noun combinations at intermediate level.

On the other hand, from the 380 verb-noun combinations produced by advanced students, 3.68% are miscollocations, 14.21% noncollocations and the remaining 82.11% are proper collocations, as displayed in Figure 16 below.

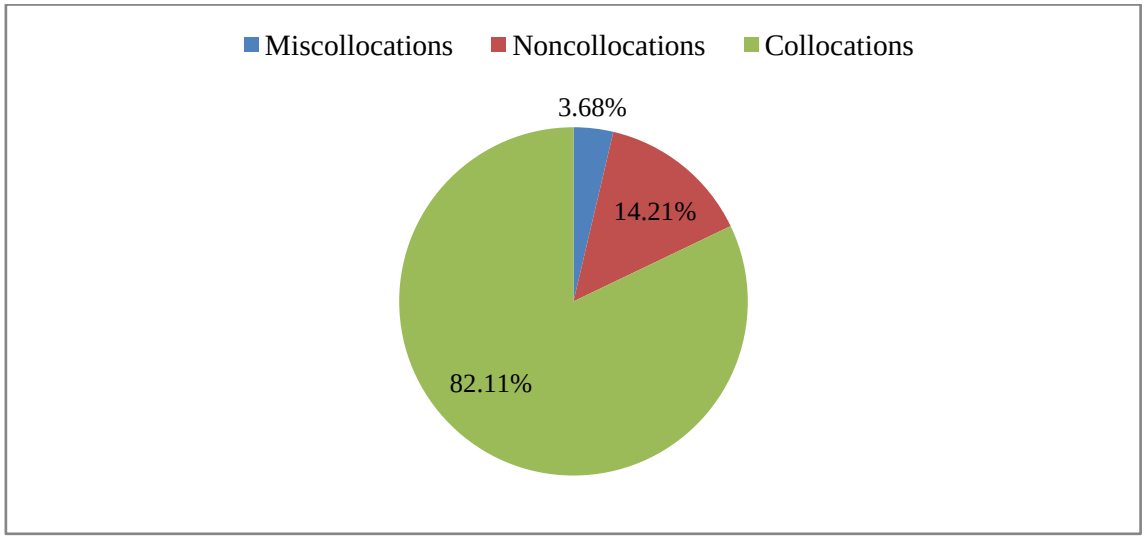

Figure 16: Verb-noun combinations at advanced level.

In conclusion, comparing the NF of the total number of verb-noun combinations produced at both proficiency levels (see Figure 14 above), we can observe that the production of collocations among intermediate students is higher than that of advanced students. As a consequence, the number of errors made by the intermediate group is also higher. It means that, although intermediate students use more collocations than advanced students in their written discourse, they also tend to make more mistakes. In sum, the advanced group produces less but more accurate combinations.

Previous studies on the subject tend to focus on one particular level, especially the advanced one, but investigations comparing two or more levels of proficiency and the production of collocations are scarce. One exception is the study carried out by Laufer and Waldman (2011) which compared the collocational production of Hebrew learners at three levels (basic, intermediate and advanced). Their results show that intermediate and advanced learners produce more collocations than basic learners, probably due to a higher degree of confidence. However, the findings show that learners who attempt to produce more collocations are likely to make errors more often. As a consequence, their study reports that the number of errors does not decrease as proficiency increases. It reports an inverse relationship between level of proficiency and correctness of collocations.

The results of this study are not in line with the findings from Laufer and Waldman (2011) since in this case learners with the highest level of proficiency do not produce more collocations than the intermediate ones. Moreover, although advanced students produce a lower number of collocations, their degree of correctness is higher. However, the results confirm the fact that learners who attempt to include more collocations in their written discourse are prone to make mistakes more frequently. A factor that could account for the differences in the results of both studies is that the L1 of the learners under study is not the same. Depending on the mother tongue of the learners, the results may vary as a consequence of the similarities and differences between the L1 and L2.

Another research that bears in mind the proficiency level of learners is the one carried out by Granger and Bestgen (2014) which examines the use of collocations by intermediate and advanced non-native writers. The study focused on bigrams, directly adjacent word pairs, which were given different scores. The results show that there are

significant differences in the use of phraseology by intermediate and advanced learners. The texts of intermediate learners are characterized by a smaller proportion of lower-frequency collocations and a higher proportion of high-frequency collocations than the texts of advanced learners. However, since the methodology used is different from the one in this study, the results obtained are not comparable.

4.2.3. Source of collocational errors

The last research question addressed in the present study is related to the sources of errors in the collocational production of Spanish learners. As has been shown in Section 4.2.1, both intermediate and advanced learners generate wrong verb-noun combinations. In this section we attempt to find an explanation for the production of miscollocations.

Data extracted from the learner corpus showed 79 occurrences of 37 different wrong verb-noun combinations produced by Spanish learners. All the miscollocations have been checked in the native corpus and dictionaries so as to verify that those combinations are not used in the written discourse of NS, with some exceptions that will be commented on below. Table 29 presents a summary of the number of miscollocations produced with each of the verbs.

	Wrong combinations	**Occurrences**
Take	9	22
Make	17	34
Have	11	23
TOTAL	**37**	**79**

Table 29: Miscollocations produced by learners.

Since there is a great number of miscollocations, only some of them will be analyzed here.[7]

[7] A complete list of wrong verb-noun combinations can be seen in Appendix II.

One of the most noticeable wrong combinations found in the learner corpus is the expression *take a coffee*, as used in example (19) below.

(19) *If you go to the bar you can´t **take a coffee** peacefully because a big dense cloud occupies the roof to the bar* (SULEC, Intermediate).

This miscollocation was produced by both intermediate and advanced learners. Intermediate students produce the combination six times (with a NF of 2.5) while advanced students generate this wrong collocation only once (with a NF of 0.48). The combination has not been considered valid since, according to the dictionaries (see Section 3.3 above), the verb that should collocate with *coffee* in that context is *have*. Using *take* instead of *have* in such context is a common error among learners which may be due to the influence of the L1, considering that in Spanish we say *tomar un café* (literally *take a coffee*).

A similar example is the wrong combination *take a drink* as in example (20) below.

(20) *When you go to xx a bar or other public places, if you are a smoker you go to the smoker's room and if you aren't a smoker you **take** your **drink** in the rooms that aren't for smokers and they are free of smoke* (SULEC, Intermediate).

As in the case of the collocation *have a coffe*e, the noun *drink* collocates with the verb *have* to create the expression *have a drink*. L1 transfer seems to be the source of error again.

Another miscollocation produced three times with the verb *take* is *take a lecture*. The sense of *lecture* in this expression is 'a serious talk to someone about their behavior', as in example (21) below.

(21) *You see childs of twelve or ten year smoke in the school, and the teachers look their, but don't **take** their **a lecture**, they look this horrible situation and don't make nothing* (SULEC, Intermediate).

This is another example of a wrong choice of verb, since according to the *Oxford Collocation Dictionary*, the verb that collocates with *lecture* to express the idea intended by the author of the essay is *give*. In this case, L1 transfer does not seem to be

the source of the error since there is no literal translation in Spanish for the expression used.

A high number of the combinations produced by learners with the verb *make* were wrong combinations. Considering the results shown in Table 29 above, this verb is the most problematic one for learners when combining it with a noun. The most recurrent miscollocation is *make (an) exam* as in example (22) below extracted from the corpus.

(22) *The University entrance examinations are something that somebody aren't agree, they think that if they pass exams of all couse, they won't want to* ***make an exam*** *which can fail their future studies* (SULEC, Intermediate).

In contrast to the previous cases mentioned, this collocation was found in the native corpus. However, there is a collocational mistake in this utterance. Learners use the expression *make an exam* in a sense different from the sense used by NS. *Make an exam* means 'create an exam', which is not the idea that the author of the text in the learner corpus tries to express. In this case, the correct expression would be *take an exam*, which means to answer the questions in an exam, not to create the exam. The source of this error could be related, as in some previous cases, to an influence of the mother tongue and a literal translation of the Spanish expression *hacer un examen*. In Spanish there is not a distinction between the person who makes the exam and the person who takes the exam, which is why learners may confuse the verbs that collocate with the English word *exam*.

Another recurrent combination is *make a question*, as in (23) below, generated by learners on five occasions.

(23) *Taking this statement as a referent there should not be any difficulty to answer the* ***question made*** (SULEC, Advanced).

According to the dictionaries of collocations consulted, the expression *make a question* is not considered as a collocation and it is not an expression used by native speakers of the language. The correct combination in Standard English would be *ask a question.* However, if we translate the expression *hacer una pregunta* literally from Spanish, the result would be *make a question* and not *ask a question*. In fact, the literal translation from the English expression (*preguntar una pregunta*) would be highly

redundant and incorrect in Spanish. This fact seems to be the main explanation for the error made by Spanish learners.

Another interesting wrong combination produced by learners is *make business*, as in (24) below.

(24) *On the one hand more people admit that they have a disagreeable problem and they want to solve their vice, although they have to go through difficult consequences like getting more weight and feeling anxious. But on the other hand more people* ***make business*** *with giving up tobacco* (SULEC, Intermediate).

The word *business* does not collocate with the verb *make* in Standard English; the correct collocation is *do business*. The most obvious explanation for this error is that both verbs (*do* and *make*) have a similar meaning in Spanish (*hacer*). As a consequence, it may be confusing for learners, given that in their mother tongue there is no distinction between those two verbs; the Spanish equivalent *hacer* is used in all the situations. An expression like *do business* is something learners have to learn by heart.

One of the most recurrent miscollocations produced with the verb *have* is *have sense*, illustrated in (25) below.

(25) *This last is an activity clearly practical, but we don't start to study it before the third year. I think it doesn't* ***have*** *much* ***sense*** *because, moreover, I consider that interpretation is even more difficult than translation, and this activity requires more practical hours* (SULEC, Intermediate).

Although *have* is a verb that, according to the dictionaries, collocates with *sense* in some cases and is found in the native corpus, the collocation *have sense* is not correct in this context. The writer of the extract tries to transmit that something is logical and, in this case, the correct combination in Standard English would be *make sense.* According to Hornby (2000), one of the meanings of that collocation is 'to be a sensible thing to do' and it is completely suitable in the context shown above.

Another interesting wrong combination is the one produced with the verb *have* and the noun *years*, as in example (26) below.

(26) *There are children who **have** 11 **years** old, and they begin this habit as a play, but it's a serious problems, and a lot of them won't go out* (SULEC, Intermediate).

There are four occurrences of this combination in the learner corpus, all of them produced by intermediate students. The expression *have x years old* does not exist in English, the grammatically correct one being *be x years old*. I consider this is an interesting mistake since one of the first expressions introduced when you start learning a language is *My name is x. I am x years old.* Therefore, it is curious that intermediate students make a mistake derived from the literal translation of the Spanish expression *Tengo x años.*

As we have already seen in the analysis of some of the wrong combinations created by learners, the influence of the mother tongue is a relevant factor in the production of verb-noun combinations in the L2, in this case English. After carrying out a more exhaustive analysis of the total number of wrong combinations, findings report that among the 37 different types of wrong combinations (out of 79 occurrences) found in the corpus with the three verbs under study, 32 are word-for-word translations of Spanish collocations that express the same meaning but should be formed with a different verb (e.g. *make emphasis* ('hacer énfasis'), *take a coffee* ('tomar un café'), *have health* ('tener salud'). The wrong choice of verb instead of the correct verb in the target language was also the most frequent type of mistake in Nesselhauf's (2003) study of collocations and free combinations with native speakers of German.

These results confirm the findings of a large number of studies in L2 acquisition which have pointed out to the vital role that the mother tongue plays in the production of collocations (see, among others, Bahns & Eldaw, 1993; Bahns, 2003; Nesselhauf, 2003, 2005; Zingraf, 2008; Fan, 2009; Laufer & Waldman, 2011; Marco, 2011). Regarding collocations, Bahns (2003) states that learners seem to rely on a 'hypothesis of transferability' (p. 61). It means that when the target collocation in the L2 has a similar structure in the L1, there is a positive transfer. However, when the two language patterns differ, there is a negative transfer that would result in an error (Ellis, 2008, p.29). Similarly, Nesselhauf (2003) refers to the congruence of the L1 and L2 structure. If a collocation can be translated word-for-word from L2 into L1, then it is congruent. If

it is not possible to translate it in such a way, the collocation is non-congruent and could lead to errors based on the L1.

Taking this into account, the present study shows evidence of a negative L1 transfer in the production of collocations. Previous studies have obtained similar results, as the one carried out by Laufer and Waldman (2001) with Hebrew students of English, in which it was also found that the main source of the error types produced by learners was a literal translation of parallel Hebrew collocations. Similarly, the study conducted by Bahns and Eldaw (1993) detected a L1 interference in the production of collocations by German advanced EFL students. Along the same lines, Fan (2009), whose study examined the collocational use of Hong Kong learners of English, points out to evidence of L2 collocation use affected by the L1. Some wrong collocations in her study were a result of a word-for-word translation of Chinese collocations and were not acceptable expressions in the English language. Finally, the findings of the studies carried out with Spanish learners of English (Zingraf, 2008; Marco, 2011) also show a negative transfer from the L1 as the main source of learners' errors in the production of collocations.

In conclusion, although the size of the learner corpus examined was limited, the results obtained confirm findings from previous studies, namely that L1 transfer may be the main source of errors in the collocational production of learners.

5. Conclusions

The main goal of this study was to discover the state-of-the-art in the collocational competence of Spanish learners of English. Two corpora have been explored in order to answer the research questions addressed in the study, the Santiago University Learner Corpus (SULEC) from the University of Santiago de Compostela in Spain, and the British National Corpus (BNC). Data obtained from the learner corpus have been analyzed and compared with material extracted from the native corpus.

The overall results of the present study provide three main findings. First of all, as previous studies show (Bahns & Eldaw, 1993; Howarth, 1996; Granger, 1998; Nesselhauf, 2003; Fan, 2009), there are certain differences in the production of collocations by NNS in comparison to NS. The results of the study show that learners underuse certain types of collocations found in the native data, as in the study carried out by Granger (1998) but, on the other hand, they tend to overuse collocations that are not so frequent in the written discourse of NS. Shih (2000) reported similar findings of an overuse of some collocations in situations where more specific expressions were required.

However, the topics of the essays in the learner corpus have proved to play an important role in the final outcomes. As shown in Section 4.2.1, the overuse of certain collocations is in most cases linked to the specific contents of the text. The topic of the essays may influence the lexical choice and the presence of some collocations that otherwise would not appear so frequently. For that reason, further research examining learner essays with a wider variety of topics should be carried out in order to investigate whether the choice of topic can be pointed out as the main cause of the overuse of certain collocations or whether there are other reasons that may account for this phenomenon.

Secondly, regarding the relationship between the level of proficiency and the collocational production, results show that intermediate students produce a higher number of collocations than advanced learners. As a consequence, the amount of errors made by the intermediate group is also higher probably because, as noted by Laufer and Waldman (2001), learners who attempt to produce more collocations are likely to make errors more often. It means that, although intermediate students use more collocations than advanced students in their written discourse, they also tend to make more mistakes.

Therefore, the advanced group produces less but more accurate combinations. The findings are not in line with the results from the study carried out by Laufer and Waldman (2011), whose study reports an inverse relationship between proficiency level and correctness of collocations, although it is interesting to note that the L1 of the learners is not the same as that of the learners in the present study. In relation to this, there is an important factor that must be considered. The proficiency level of the Spanish learners in the learner corpus is established according to their age, as mentioned above in Section 3.1.1. I believe that this classification is not completely reliable since being older does not necessarily mean having a better command of English. For that reason, I reckon it should be necessary to test learners' level of competence before establishing those categories in order to be able to get more reliable data.

Thirdly, the analysis of collocational errors produced by Spanish learners has shown that L1 influence plays a major role in the production of collocations in the L2. Thus, it confirms the findings from several previous studies (Bahns & Eldaw, 1993; Bahns, 2003; Nesselhauf, 2003, 2005; Zingraf, 2008; Fan, 2009; Laufer & Waldman, 2011; Marco, 2011) that point out to L1 transfer as the main source of problems for learners in the production of collocations.

Although the main goals of the study have been fulfilled, there are some limitations that should be considered. Some of them have been already mentioned, as the choice of the topic and the proficiency level of learners according to their age. As another suggestion for further research, the results in this work should be verified by statistical analysis of the data obtained, especially to check if the difference in the production of collocations between NNS and NS is significant. Such an analysis was beyond the scope of the present research, but will be considered in the future.

Another interesting suggestion for further research is the categorization of the collocations extracted into a cline of idiomaticity according to their degree of restriction, so as to examine whether there is any relationship between the degree of restriction of a combination and the mistakes made by the learners, as investigated by Nesselhauf (2003). She concluded that "collocations with a low degree of restriction are the most difficult kind of combination for the learners" (p. 234). It would be interesting to check whether this is also the case with Spanish learners of English.

To conclude, the results in this study shed new light on the collocational production of Spanish learners. Nevertheless, more studies examining different learner corpora should be carried out in order to obtain more reliable results. Although the size of the learner corpus examined (455,000 in the written section) is not small in comparison to other corpus of Spanish learners (e.g. Spanish sub-corpora of International Corpus of Learner English, ICLE), the results cannot be generalized to the entire population of Spanish learners of English. The sample of this study corresponds to one corpus of a specific university and different results could be offered by examining other corpora of Spanish learners. In addition, the university students participating in this project belong to an English Philology degree, which means that students taking other degrees could have provided different outcomes. As a consequence, further research should be carried out comparing the results from different Spanish learners' corpora to find out whether the data in the present work are confirmed.

6. References

Aisenstadt, E. (1981). Restricted collocations in English Lexicology and Lexicography. *Review of Applied Linguistics, 53*, 53-62.

Bahns, J. (1993). Lexical collocations: A contrastive view. *ELT Journal, 47(1)*, 56-63. Retrieved July 13, 2015, from http://dx.doi.org/10.1093/elt/47.1.56

Bahns, J. & Eldaw, M. (1993). Should we teach ESL students collocations? *System, 21(1)*, 101-114.

Barfield, A. (2003). *Collocation recognition and production: Research insights*. Tokyo: Chuo University.

Benson, M., Benson, E., & Ilson, R. (1985). The structure of the collocational dictionary. *International Journal of Lexicography, 2*, 1-14. Retrieved July 15, 2015, from http://dx.doi.org/10.1093/ijl/2.1.1

Benson, M., Benson, E., & Ilson, R. (1986). *The BBI combinatory dictionary of English: A guide to word combinations*. Amsterdam/Philadelphia: John Benjamins. Retrieved July 15, 2015, from http://bbidict.narod.ru/

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. London: Longman.

Carter, R. (1998). *Vocabulary: Applied linguistic perspectives* (2nd ed.) London: Routledge.

Conzett, J. (2000). Integrating collocation into a reading and writing course. In M. Lewis (Ed.), *Teaching collocation: Further developments in the lexical approach* (pp. 70-86). Hove: Language Teaching Publications.

Cowie, A. P. (1981). The treatment of collocations and idioms in learners' dictionaries. *Applied Linguistics, 2(3)*, 223-235.

Davies, Mark. (2004). The British National Corpus BYU-BNC. Based on the British National Corpus from Oxford University Press. Available online at http://corpus.byu.edu/bnc/.

Fan, M. (2009) An exploratory study of collocational use by ESL students – a task based approach. *System, 37(1),* 110-123.

Firth, J. R. (1957). A synopsis of linguistic theory, 1930-55. (Reprinted in *Selected papers of J. R. Firth,* pp. 168-205, by F. R. Palmer, Ed., 1968, London: Longmans)

Granger, S. (1998). Prefabricated patterns in advanced ESL writing: Collocations and formulae. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis and applications* (pp. 145-160). Oxford: Oxford University Press.

Granger, S., & Bestgen, Y. (2014). The use of collocations by intermediate vs. advanced non-native writers: A bigram-based study. *International Review of Applied Linguistics in Language Teaching*, *52(3)*, 229-252.

Gyllstad, H. (2007). *Testing English collocations: Developing receptive tests for use with advanced Swedish learners*. Lund: Lund University.

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.

Hill, J. (2000). Revising priorities: From grammatical failure to collocational success. In M. Lewis (Ed.), *Teaching collocation: Further developments in the lexical approach* (pp. 47-69). Hove: Language Teaching Publications.

Hornby, A. S. (2000) *Oxford advanced learner's dictionary* (8thed.). Oxford: Oxford University Press.

Howarth, P. (1996). *Phraseology in English academic writing: Some implications for language learning and dictionary making*. Tübingen: Narr

Howarth, P. (1998). Phraseology and second language proficiency. *Applied Linguistics, 19 (1),* 24-44.

Kurosaki, S. (2013). *An analysis of the knowledge and use of English collocations by French and Japanese learners* (Unpublished doctoral dissertation). University of London Institute in Paris, Paris.

Lado, R. (1957). *Linguistics across cultures: Applied linguistics for language teachers.* Ann Arbor: University of Michigan Press.

Laufer, B. & Waldman, T. (2011). Verb-noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning, 61(2),* 647-672.

Lewis, M. (2000). Learning in the lexical approach. In M. Lewis (Ed.), *Teaching collocation: Further developments in the lexical approach* (pp.126-153). Hove: Language Teaching Publications.

Li, J., & Schmitt, N. (2010). The development of collocation use in academic texts by advanced L2 learners: A multiple case-study approach. In D. Wood (Ed.), *Perspectives on formulaic language: Acquisition and communication* (pp. 23-46). London/New York: Continuum.

Lukač, M., & Takač, V. (2013). How word choice matters: An analysis of adjective-noun collocations in a corpus of learner essays. *Jezikoslovlje 14(2-3),* 385-402.

Marco, M. J. L. (2011). Exploring atypical verb+noun combinations in learner technical writing. *International Journal of English Studies, 11(2),* 77–95.

Nattinger, J. R., &DeCarrico, J. S. (1992). *Lexical phrases and language teaching*. Oxford: Oxford University Press.

Nesselhauf, N. (2003). Transfer at the locutional level: An investigation of German-speaking and French-speaking learners of English. In C. Tschichold (Ed.), *English core linguistics. Essays in honour of D.J. Allerton* (pp. 269-286). Bern: Peter Lang.

Nesselhauf, N. (2004). What are collocations? In D. Allerton, N. Nesselhauf & P. Skandera (Eds.). *Phraseological units: Basic concepts and their application* (pp. 1-21). Basel: Schwabe.

Nesselhauf, N. (2005). *Collocations in a learner corpus*. Amsterdam/Philadelphia: John Benjamins.

Oxford online collocation dictionary. Available at http://oxforddictionary.so8848.com/.

Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R.W. Schmidt (Eds.), *Language and communication* (pp. 191-226). New York: Longman.

Schmitt, N. (2004). *Formulaic sequences: Acquisition, processing and use*. Amsterdam/Philadelphia: John Benjamins.

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. London: Palgrave Macmillan.

Shih, R. (2000). Collocation deficiency in a learner corpus of English: From an overuse perspective. *PACLIC 14 Proceedings*, 281-288.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Siyanova, A., & Schmitt, N. (2008). L2 Learner production and processing of collocation: A multi-study perspective. *The Canadian Modern Language Review 64(3)*, 429-458.

Wray, A. (2002). *Formulaic language and the lexicon.* Cambridge: Cambridge University Press.

Zinkgräf, M. (2008). V+N miscollocations in the written production of university level students. *ELIA (Estudios de Linguistica Inglesa Aplicada), 8*, 91-116.

Appendix I

Verb-noun combinations (SULEC)

VERB	NOUN	FREQUENCY		
		INTERMEDIATE	ADVANCED	TOTAL
Take	*Place*	12	4	**16**
Take	*Care*	7	7	**14**
Take	*Decisions*	9	3	**12**
Take	*Decision*	7	5	**12**
Take	*Measures*	5	6	**11**
Take	*Drugs*	6	2	**8**
Take	*Cigarette*	4	3	**7**
Take	*Coffee*	6	1	**7**
Take	*Time*	4	1	**5**
Take	*Risk*	1	3	**4**
Take	*Part*	2	1	**3**
Take	*Degree*	-	3	**3**
Take	*Rules*	3	-	**3**
Take	*Lecture*	3	-	**3**
Take	*Control*	2	-	**2**
Take	*Steps*	2	-	**2**
Take	*Action*	2	-	**2**
Take	*Plane*	-	2	**2**
Take	*Hours*	1	1	**2**
Take	*Option*	2	-	**2**
Take	*Air*	2	-	**2**
Take	*Illness*	1	1	**2**
Take	*Drink*	1	1	**2**
Take	*Exam*	1	-	**1**
Take	*Advice*	-	1	**1**
Take	*Advantage*	-	1	**1**
Take	*Bus*	-	1	**1**
Take	*Exercise*	1	-	**1**
Take	*Bottle*	1	-	**1**
Take	*Profits*	1	-	**1**
Take	*Look*	1	-	**1**
Take	*Fun*	1	-	**1**
Take	*Change*	-	1	**1**
TOTAL		88	48	**136**

VERB	NOUN	FREQUENCY		
		INTERMEDIATE	ADVANCED	TOTAL
Make	*Laws*	11	2	**13**
Make	*Effort*	7	3	**10**
Make	*Money*	7	3	**10**
Make	*Law*	4	2	**6**
Make	*Exam*	6	-	**6**
Make	*Difference*	4	2	**6**
Make	*Mistake*	2	4	**6**
Make	*Question*	4	1	**5**
Make	*Exams*	4	-	**4**
Make	*Changes*	4	-	**4**
Make	*Cigarette*	4	-	**4**
Make	*Family*	2	2	**4**
Make	*Decisions*	1	2	**3**
Make	*Decision*	2	1	**3**
Make	*Love*	2	1	**3**
Make	*Use*	1	2	**3**
Make	*Demonstration*	1	2	**3**
Make	*Sense*	2	-	**2**
Make	*Comparison*	2	-	**2**
Make	*Joke*	2	-	**2**
Make	*Attempt*	1	1	**2**
Make	*Distinction*	-	2	**2**
Make	*Exercise*	2	-	**2**
Make	*Pressure*	2	-	**2**
Make	*Emphasis*	-	2	**2**
Make	*Noise*	1	-	**1**
Make	*Assertion*	1	-	**1**
Make	*Selection*	1	-	**1**
Make	*Proposal*	1	-	**1**
Make	*Discovery*	1	-	**1**
Make	*Complaint*	-	1	**1**
Make	*Request*	1	-	**1**
Make	*Progress*	1	-	**1**
Make	*Statement*	-	1	**1**
Make	*Diet*	1	-	**1**

Make	*Explanation*	1	-	**1**
Make	*Part*	1	-	**1**
Make	*Investigation*	1	-	**1**
Make	*Fire*	1	-	**1**
Make	*Business*	1	-	**1**
Make	*Pain*	1	-	**1**
Make	*Opinion*	1	-	**1**
Make	*Harm*	-	1	**1**
Make	*Damage*	-	1	**1**
TOTAL		92	36	**128**

VERB	**NOUN**	**FREQUENCY**		
		INTERMEDIATE	**ADVANCED**	**TOTAL**
Have	Rights	112	55	**167**
Have	Right	81	34	**115**
Have	Problems	55	8	**63**
Have	Children	40	19	**59**
Have	Money	25	29	**54**
Have	Problem	26	8	**34**
Have	Opportunity	18	7	**25**
Have	Opinion	19	6	**25**
Have	Freedom	19	1	**20**
Have	Knowledge	11	7	**18**
Have	Opportunities	12	5	**17**
Have	Friends	13	3	**16**
Have	Baby	12	3	**15**
Have	Possibilities	11	3	**14**
Have	Illness	12	2	**14**
Have	Advantages	11	3	**14**
Have	Experience	9	5	**14**
Have	Importance	6	8	**14**
Have	Respect	12	1	**13**
Have	Cancer	11	1	**12**
Have	Job	9	3	**12**

Have	Consequences	6	5	**11**
Have	Possibility	5	5	**10**
Have	Habit	8	2	**10**
Have	Coffee	6	4	**10**
Have	Reasons	7	2	**9**
Have	Idea	6	3	**9**
Have	Disadvantages	8	1	**9**
Have	Value	6	2	**8**
Have	Time	6	2	**8**
Have	Effect	4	4	**8**
Have	Chance	5	3	**8**
Have	Difficulties	5	3	**8**
Have	Feelings	5	3	**8**
Have	Ideas	7	-	**7**
Have	Disease	4	3	**7**
Have	Option	4	3	**7**
Have	Question	3	3	**6**
Have	Role	4	2	**6**
Have	Family	4	2	**6**
Have	Contact	4	1	**5**
Have	Responsabilities	3	2	**5**
Have	Doubt	-	4	**4**
Have	Sense	4	-	**4**
Have	Fun	2	2	**4**
Have	Years	4	-	**4**
Have	Accident	3	1	**4**
Have	Addiction	3	1	**4**
Have	Dinner	3	1	**4**
Have	Reason	2	1	**3**
Have	Lunch	-	3	**3**
Have	Influence	3	-	**3**
Have	Choice	1	2	**3**
Have	Prejudices	3	-	**3**
Have	Sex	3	-	**3**
Have	Probabilities	2	1	**3**
Have	Alternative	2	1	**3**
Have	Punishment	3	-	**3**

Have	War	-	2	**2**
Have	Impression	-	2	**2**
Have	Impact	-	2	**2**
Have	Hope	2	-	**2**
Have	Access	2	-	**2**
Have	Guilt	2	-	**2**
Have	Health	2	-	**2**
Have	Drugs	1	1	**2**
Have	Break	1	1	**2**
Have	Consideration	2	-	**2**
Have	Validity	-	1	**1**
Have	Status	-	1	**1**
Have	Beer	-	1	**1**
Have	Chat	-	1	**1**
Have	Belief	-	1	**1**
Have	Principles	1	-	**1**
Have	Care	1	-	**1**
Have	Necessity	1	-	**1**
Have	Differences	1	-	**1**
Have	Blame	1	-	**1**
Have	Control	1	-	**1**
Have	Risk	1	-	**1**
Have	Dream	1	-	**1**
Have	Trouble	1	-	**1**
Have	Breakfast	1	-	**1**
Have	Frontiers	1	-	**1**
Have	Relevance	1	-	**1**
Have	Allergy	1	-	**1**
TOTAL		697	296	**993**

Appendix II

Wrong verb-noun combinations (SULEC)

VERB	NOUN	OCCURRENCES		
		Intermediate	**Advanced**	**Total**
Take	*Coffee*	6	1	7
Take	*Rules*	3	-	3
Take	*Lecture*	3	-	3
Take	*Air*	2	-	2
Take	*Illness*	1	1	2
Take	*Drink*	1	1	2
Take	*Profits*	1	-	1
Take	*Fun*	1	-	1
Take	*Change*	-	1	1
Make	*Exam*	6	-	6
Make	*Question*	4	1	5
Make	*Exams*	4	-	4
Make	*Demonstration*	1	2	3
Make	*Exercise*	2	-	2
Make	*Pressure*	2	-	2
Make	*Emphasis*	-	2	2
Make	*Diet*	1	-	1
Make	*Explanation*	1	-	1
Make	*Part*	1	-	1
Make	*Investigation*	1	-	1
Make	*Fire*	1	-	1
Make	*Business*	1	-	1
Make	*Pain*	1	-	1
Make	*Opinion*	1	-	1
Make	*Harm*	-	1	1
Make	*Damage*	-	1	1
Have	*Sense*	4	-	4
Have	*Years*	4	-	4
Have	*Punishment*	3	-	3
Have	*War*	-	2	2
Have	*Guilt*	2	-	2

Have	*Health*	2	-	2
Have	*Drugs*	1	1	2
Have	*Care*	1	-	1
Have	*Blame*	1	-	1
Have	*Control*	1	-	1
Have	*Risk*	1	-	1
TOTAL		**65**	**14**	**79**