

TAKE + NOUN SEQUENCES IN NATIVE AND LEARNER WRITTEN DATA¹

ELSA M. GONZÁLEZ ÁLVAREZ

SUSANA M. DOVAL SUÁREZ

UNIVERSIDAD DE SANTIAGO DE COMPOSTELA

Abstract. Multi-word sequences have been shown to pose important problems for learners even at the most advanced levels (Nesselhauf 2003, 2005). In this article we compare the use of the high-frequency verb *take* by both native speakers (NS) and advanced Spanish-speaking learners of English (NNS) in three types of multiword sequences (free combinations, collocations and idioms), although the focus will be mainly on collocations. The data were drawn from the Spanish subsection of the *International Corpus of Learner English* and the *Louvain Corpus of Native Speaker Essays*. Following the framework of Contrastive Interlanguage Analysis (Granger *et al.* 2002), we studied the learners' linguistic behaviour from the perspective of what combinations they use *significantly more* or *significantly less* than native speakers. The results show that learners significantly underuse collocations and free combinations, but significantly overuse idioms.

Key words: *Collocations, Idioms, Free Combinations, Advanced Learner, Contrastive Interlanguage Analysis.*

1. Introduction

1.1. Literature review

There is general agreement that multi-word sequences such as collocations and idioms constitute an important part of native-speaker communicative competence (Howarth 1998, Nesselhauf 2003, Keshavarz and Salimi 2007) “and make up a large part of any discourse” (Conklin and Schmitt 2008: 74). As Sinclair (1991) observed, language users resort to the large number of semi-preconstructed phrases available to them, that is, the “idiom principle”, far more often than to the “open principle”, or “slot and filler” model, by which virtually any word which satisfies local constraints can occur at each slot (Sinclair 1991: 110). Conklin and Schmitt (2008) provide sociofunctional and psycholinguistic explanations for the pervasiveness of formulaic sequences which point to their pragmatic value and processing efficiency. These researchers claim that “a significant processing advantage was observed for formulaic sequences over nonformulaic language (...) for both L1 and L2 English speakers” (Conklin and Schmitt 2008: 85).

Similarly, collocational competence seems to be a crucial component in SLA, not only because it reduces the cognitive challenge that L2 production and processing poses on learners (Conklin and Schmitt 2008), but also because failure to use collocations accurately adds to learners' foreign-soundingness (Hsu and Chiu 2008). As Nesselhauf (2003: 223) puts it, “collocations not only enhance accuracy but also fluency”. In fact, a number of corpus-based studies have highlighted that these constructions pose important problems for learners even at the most advanced levels (Källkvist 1995, Granger 1998, Lorenz 1999, Nesselhauf 2003, 2005). It has been also suggested that this is especially true of high-frequency words, due to their semantic opacity and restricted collocability. A good illustration of this phenomenon is found in the delexicalised uses of high-frequency verbs, i.e. those uses

¹ This research has been funded by the Xunta de Galicia (INCITE09 204 155 PR) and the Spanish Ministry of Innovation and Science (FFI2010-19380). These grants are hereby gratefully acknowledged.

REVISTA ELECTRÓNICA DE LINGÜÍSTICA APLICADA (ISSN 1885-9089)

2011, Número 10, páginas 55-68

Recibido: 13/12/2010

Aceptación comunicada: 14/03/2011

occurring mainly in phraseological patterns, which present serious problems for learners, especially in production (de Cock and Granger 2004, Gouverneur 2008).

However, in spite of the importance of phraseology in language use and L2 acquisition, studies that provide a detailed description of learners' and native speakers' phraseological performance do not precisely abound (Howarth 1998, Conklin and Schmitt 2008), and have yielded contradictory results. Thus, whereas Granger (1998: 146) formulates the initial hypothesis that learners would resort to conventionalized language more rarely than their NS counterparts, since the use of such language is universally presented as native-like, Conklin and Schmitt (2008: 76), however, note that evidence has been found indicating that non-natives rely on formulaic language a great deal in their efforts to produce fluent speech.

Regarding collocations, in spite of their widely acknowledged importance for learners, there are not many studies that analyse non-native speakers' use of collocations. Nesselhauf (2003: 224) points out that such studies are rare and unsatisfactory either because of the data-elicitation methods used or because of the vague definition of the concept of collocation.

A number of studies (Zhang 1993, Sung 2003, Keshavarz and Salimi 2007) have revealed a significant positive correlation between learners' knowledge of lexical collocations and their language proficiency. L2 learners, however, appear to have a better receptive than productive knowledge of collocations, as Hsu and Chiu (2008: 184-185) found in a study which showed that such positive correlation did not exist between learners' use of collocations and speaking proficiency, nor between subjects' knowledge and use of lexical collocations. Similarly, Zhang and Chen (2006) designed a test to examine the receptive and productive knowledge of adjective + noun collocations by three groups at different proficiency levels. The results show that, although differences between and within groups were found, in general the subjects did not apparently have a good command of these collocations, and that transformation from receptive knowledge into productive skills seemed to be a slow process.

In other studies of language proficiency, different language measurements have yielded contradictory results. Howarth (1998) administered a language test to L2 English learners and found no correlation between scores in the test and knowledge of collocations, whereas Keshavarz and Salimi (2007) found positive correlations between the scores on collocation tests and scores on the cloze tests, which strongly suggests that learners' collocational competence and proficiency level are positively associated. Howarth (1998), however, admits that the language test was not a valid measure of proficiency.

Different definitions have been proposed for the term collocation by researchers using different criteria to delimit them from other types of word combinations. A first distinction can be drawn between phraseological and statistical criteria with some authors using both. According to Koya (2004), phraseological criteria are subjective and are concerned with collocational restriction and semantic opacity whereas objective criteria are related to statistics and are frequency-based, that is, are related to the co-occurrence of words in a certain span (Nesselhauf 2003: 224).

It is generally agreed by authors within phraseological approaches that it is difficult to draw a line between types of combinations that are often thought of as placed along a continuum of restriction on substitutability. Within this phraseological tradition, Gouverneur (2008: 232) follows Cowie (1998) who defined restricted collocations as "word combinations in which some substitution is possible, but with some arbitrary limitations on substitution; in which at least one element has a non-literal meaning, and at least one element is used in its literal sense; and the whole combination is transparent". Similarly, Van Roey (1990: 46) observes that those arbitrary limitations on substitution are on the level of usage rather than of syntax or meaning.

Also within the phraseological tradition, Howarth (1998) proposes a framework for the study of collocations adapted from the continuum model developed by Soviet phraseologists and discussed, among others, by Cowie (1988). Three major categories mark end and middle points along the axes of semantic specialization and substitutability of constituent elements and represent three main levels of restrictedness: *free combinations*, with the highest degree of semantic transparency and substitutability, *idioms*, with the lowest degree of both, and, between them, *restricted collocations* (Howarth 1998). Idioms are further subdivided into *figurative idioms*, which “have metaphorical meanings in terms of the whole and have a current literal interpretation” (Howarth 1998: 28) and *pure idioms*, whose meaning cannot be predicted from the individual meanings of its components. Similarly, the central category of restricted collocations can be broken into subdivisions depending on how strictly the criterion of commutability is applied. However, Howarth (1998:28) admits that “there are clearly problems in making this method of categorization reliable. The difficulty lies in finding an authority for deciding on what substitutions are ‘permitted’”.

Following Cowie’s (1988) definition, Nesselhauf (2003) refines the notion of “restricted sense” which she believes helps to delimit the three main types of word combinations and to clarify the distinction between arbitrary and semantically motivated restriction. If at least one of two criteria applies, a sense of a verb (or a noun) is considered “restricted”: 1) if it “is so specific that it only allows its combination with a small set of nouns (verbs)” and/or 2) it “cannot be used in this sense with all nouns (verbs) that are syntactically and semantically possible ” (Nesselhauf 2003: 226).

Free combinations are defined, then, as combinations in which both the verb and the noun are used in unrestricted senses. Combinations where the noun is used in an unrestricted sense, but the verb is used in a restricted sense are classified as collocations (i.e. *take a picture*). Finally, idioms are those word combinations where both the elements are used in a restricted sense.

Nesselhauf (2003: 227) admits that the line between the three types is still difficult to draw and not rigid and there still remains the problem of determining whether a combination “exists” in a language or not.

The criteria just described are, as Koya (2004) observes, a matter of degree and to some extent subjective. More objective are criteria related to statistics, which identify collocations by means of procedures which establish the frequency of co-occurrence of the elements in a collocation or the strength of association between the collocate and the node (Dayrell 2007).

In an attempt to overcome the drawbacks of purely statistical criteria, some authors have used a combination of linguistic analysis and statistical measures. Based on a linguistic analysis of collocations within the framework of the Meaning-Text theory, Orliac (2008) extracted V+N combinations from a corpus of computer science texts and subsequently used statistical measures to distinguish true collocations from free combinations (Orliac 2008).

As regards the methods of data collection, two groups of studies can be identified in SLA for the analysis of collocations: corpus-based and experimental. Experimental studies have designed procedures to elicit collocations, test receptive knowledge or judge the acceptability of accepted collocations (Gabrys-Biskup 1992, Bahns 1993, Bahns and Eldaw 1993, Zhang and Chen 2006).

One drawback of this approach, as Howarth (1998: 32) observes, is “the difficulty of establishing the validity of any predefined list of target collocations, since, as many EFL teachers might agree, this component of a learner's linguistic competence is one of the least predictable”. Corpus studies could overcome this drawback, since, as Howarth (1998) observes, they examine complete texts focusing on their phraseological features, and not on a

predetermined set of collocations. He mentions Granger's work as an example of such an approach (Howarth 1998).

A number of studies concentrate on lexical collocations, particularly on verb + noun constructions. Howarth (1998) compared the frequency of the three main types of constructions in native and non native speakers' written production and found that 62% could be classified as free collocations, 33% as restricted collocations and 5% as idioms in the first group. Comparable figures were obtained by Cowie (1991, 1992) who found that between 37.5% and 46% of all the verb + direct object constructions were classified as restricted collocations or idioms. Howarth (1998) compares the proportion of the two categories regarded as conventional (restricted collocations and idioms) in the two groups: 38% in NS and 25% in NNS (24% restricted collocation and 1% idiom). The author interprets this difference as an indicative of "either a generally lower level of knowledge of collocations, or a lack of awareness of how to deploy them appropriately, or both (a finding confirmed by Granger forthcoming)" (Howarth 1998: 36). A fourth category, deviant constructions, was added to classify non native speakers' data, which accounted for 6% of the constructions identified.

1.2. Statement of purpose

The aim of this paper is to compare the use of the high-frequency verb *take* by both native speakers and advanced Spanish-speaking learners in three types of multiword sequences (free combinations, collocations and idioms), although the focus will be mainly on collocations. The learner's linguistic behaviour will be analysed from the perspective of what combinations they use significantly more or significantly less than a NS (Contrastive Interlanguage Analysis, Granger *et al.* 2002). More specifically, the study aims at answering the following questions:

1. Do advanced EFL learners and NS use the different types of *take*-combinations with the same or similar frequency?
2. What are the reasons for the possible differences between the two groups?
3. Is there any relationship between the number and type of mistakes made by the learner and the degree of restriction of a combination?
4. Similarly, is there any relationship between the number and type of mistakes made by the learner and the different senses of the verb *take*?

The verb *take* was selected for the study because, although typically classified as an "easy" verb, it presents difficulties to the learner especially as regards phraseology, as well as other frequent verbs (Gouverneur 2008):

"Compared to native speakers, learners not only tend to mainly overuse these verbs (Altenberg and Granger 2001, Kaszubski 2000) but they also misuse them to a great extent (Nesselhauf 2004). Whilst the core meanings of these verbs usually seem to be mastered, their delexicalised uses, occurring mainly in phraseological patterns, have shown to remain a stumbling block to native-like proficiency" (Gouverneur 2008: 223-224).

2. Method

This study uses a Corpus Linguistics methodology. The data were taken from two comparable corpora: The NNS corpus used was the Spanish component of the *International*

Corpus of Learner English (SPICLE). The Spanish subcorpus comprises 251 essays, with a total of 200,376 words.

As stated in Granger, Dagneaux and Meunier (2002:13), “all the learners submitting an essay to the ICLE shared the following characteristics: age, learning context and proficiency level”. This means that: 1) the subjects are usually in their twenties; 2) they have learned English in a non-English speaking country and 3) all the ICLE subjects are university undergraduates in English, usually in their third or fourth year, with the proficiency level ranging from higher intermediate to advanced.

The NS control corpus used was the LOCNESS (the Louvain Corpus of Native English Essays), a 300,000-word corpus, including 149,574 words of argumentative essays written by American university students, 18,826 words of literary-mixed essays written by American university students, 59,568 words of argumentative and literary essays written by British university students, and 60,209 words of British A-level argumentative essays. It has the advantage of being directly comparable to the ICLE, but, in order to make it comparable to the Spanish subsection of the ICLE, we limited our study to a subsection of the LOCNESS, a sample of 322 essays (totalling 227,968 words) which excludes essays produced by A-level students.

The first step in our research involved the extraction of all the instances of the lemma *take* in both corpora using Wordsmith Tools. The selection of *take* + noun combination was carried out manually, as well as sorting all the instances into the categories and meanings previously defined. For the semantic classification, we followed Gouverneur’s (2008) categorisation of the senses of *take*, done on the basis of previous studies, and on categories from commonly used learner’s dictionaries. Her classification is reproduced below:

Meanings and patterns	Examples
1. Move	<i>Can you take us to the airport?</i>
2. Eat or drink	<i>They take drugs; do you take milk?</i>
3. Phrasals	<i>Take off</i>
4. Need	<i>Take the time to read it; It takes a year to...</i>
5. Delexical uses	<i>To take action</i>
6. Think of in a certain way	<i>Take it easy</i>
7. Accept	<i>Staff will be available to take your enquiry</i>
8. Transport	<i>We took the ferry</i>
9. Other uses	

The categorisation into different types of combinations was done based on Nesselhauf’s (2003) work, who classified verb-noun combinations as to their degree of restriction into free combination, restricted collocations and idioms and as to their degree of acceptability into correct or wrong. Assigning examples to pre-conceived categories is an extremely difficult task and no classification is uncontroversial or without limitations. In order to overcome this difficulty, two dictionaries, the Oxford Collocations Dictionary for Students of English and the Cambridge Advanced Learners’ Dictionary, were used to assign the examples found to the three categories mentioned above.

As for the degree of acceptability (Correct - Wrong), a categorisation of four types of errors was established:

A- wrong choice of verb

- (1) *take the role [spm06002] => play the role

(2) *make advantage => take advantage (invented example)

B- wrong choice of noun

(3) *take a sight [spm07010] => take a look

C- meaning

(4) * they should take care of their words [spm03038] => be careful with their words

D- pre- or post-modification

(5) *take a look to [spm10002] => take a look at

An additional limitation in the scope of this paper, regarding types of mistakes, should be mentioned. Since only *take* + noun combinations are considered here, there is no way to know if another combination with a different verb is used in place of the “correct” combination. Therefore, no statistical analyses were carried out to compare the significance of the different types of mistakes since the data for the first category would not be real. The categorisation was used only for descriptive purposes. Further research in which all verb + noun combinations were analysed would be needed to solve this problem.

Therefore, wrong combinations were identified only when *take* was used instead of another (correct) verb, as in example (1) above, or when a wrong noun was used (3), and not when another verb was used instead of *take* (2). A different type of mistake was also identified, and involved the use of a correct combination but with a meaning that was incorrect in the context (4). Combinations where non-lexical elements such as prepositions differed from NS use were also classified as errors (5). Grammar (i.e. tense) or spelling errors were overlooked.

The last step involved carrying out statistical analyses in order to determine what combinations the learner uses significantly more or significantly less than a native speaker. The methodology used to describe quantitative differences between NS and NNS is *Contrastive Interlanguage Analysis* or CIA (Granger 1996, Granger *et al.* 2002), which involves two major types of comparison: 1) NL *vs.* Interlanguage (IL) comparisons, whose aim is to “uncover the features of non-nativeness of learner language. At all levels of proficiency, but especially at the *most advanced* ones, these features will not only involve plain errors, but differences in the frequency of use of certain words, phrases or structures, some being overused, others underused” (Leech 1998: 13; the emphasis is ours); 2) IL *vs.* IL. This study focuses on the first type of comparison.

In line with other CIA studies, we view the labels *overuse* and *underuse* as purely descriptive, which means that they do not necessarily imply wrong usage. Although the study presented here is predominantly quantitative, some more qualitative insights into the use of *collocations* are also provided in order to highlight the importance of qualitative considerations when dealing with the acquisition of *take* + noun combinations. Qualitative analyses were also performed to try to provide explanations for the numerical data obtained.

3. Analysis and results

3.1 The use of Free Combinations (F), Collocations (C) and Idioms (I) by NS and NNS

The total number of *take*-combinations was analysed for both corpora. Table 1 shows the total number of combinations and the results for the different types.²

	<i>NS</i>	<i>%</i>	<i>NNS</i>	<i>%</i>	TOTAL
C	318	61.6	175 -**	59.1	493
F	184	35.6	82 -**	27.7	266
I	14	2.7	39 +**	13.1	53
TOTAL	516		296 - **		812

Table 1: Raw frequencies of *take* + noun combinations.

The results indicate the existence of significant statistical differences in the total number of combinations in both groups ($t=2.565$; $p=0.011$), that is, NS use significantly more *take*-combinations than NNS. The data also show that learners significantly underuse collocations ($t=2.111$; $p=0.035$) and free combinations ($t=3.743$; $p=0.000$), but significantly overuse idioms ($t=4.082$; $p=0.000$). However, the same order was found in both groups, although with different relative frequencies: collocations are the most frequent type, followed by free combinations; idioms are the least frequent type.

3.2 Relationship between the degree of restriction of a combination and learner mistakes

Following Nesselhauf (2003), we also wanted to find out if there was any relationship between the degree of restriction of a combination and the mistakes made by the learners.

	<i>Total</i>	<i>errors</i>	<i>%</i>
C	175	14	8%
F	82	8	10%
I	39	2	5%
Total	296	24	11%

Table 2: Relationship between combination type and number of errors in ICLE.

As we can see in table 2, the highest rate of mistakes is found in free combinations, where 10% of the combinations found in the NNS corpus had some type of mistake. These are closely followed by collocations (8%), whereas the lowest percentage occurs in idioms (5%). Although Nesselhauf (2003) obtained rather different results, we can use her explanation to interpret our data: learners produce fewer mistakes in combinations which are acquired and produced as wholes, whereas free combinations may often be “too creative” (Nesselhauf 2003: 233). She found the highest percentage of mistakes in collocations with a medium degree of restriction (33%) and the lowest in collocations with a lot of restriction (18%), with idioms and free combinations in between (23%). She concludes that “the degree of restriction does not have a major influence on the types and amount of mistakes learners make, except that collocations with a low degree of restriction are the most difficult kind of combination for the learners” (Nesselhauf 2003: 234).

We also analysed the different types of mistakes and whether different types of combinations favoured particular types of mistakes. The results are shown in table 3.

² Significant levels of underuse or overuse on the part of the learners are indicated by a plus or a minus sign followed by a double asterisk.

COMB. TYPE	Total errors	%	verb	%	noun	%	meaning	%	pre-modif	%	post-modif	%
C	14	100	5	36	1	7	2	14	1	7	5	36
F	8	100	7	88		0	1	13		0		0
I	2	100		0	1	50		0	1	50		0
Total	24	100	12	50	2	8	3	13	2	8	5	21

Table 3: Relationship between error types and combination types in NNS.

Half of the mistakes found in all the categories consist in *wrong choice of the verb*, that is, the use of *take* instead of the target verb. This is the most frequent type of mistake in collocations (36%), together with *postmodification* (often preposition) errors, and in free combinations (88%). This was also the most frequent type of mistake in Nesselhauf's (2003) study both for collocations and free combinations.

3.3 NNS overuse of idioms

An analysis of the different types of idioms was carried out in order to explain the overuse of idioms on the part of NNS (table 4), which turned out to be due to the high frequency of a single type (*take something into account*). This type is, in turn, also the most frequent one in NS data, what could explain learners' overuse of this idiomatic expression.

NS	NNS	BOTH
Take sth into account (11) Take sth for granted (3)	take sth into account (31) take sth for granted (3) *take account (3) take issue with (1) *take sth into analysis (1)	Take sth into account Take sth for granted

Table 4: Idioms.

3.4. Collocation types and tokens in NS and NNS

The number of types and tokens of *take*-collocations in the two corpora were analysed and the results are shown in tables 5 and 6 respectively. The t-test revealed a statistically significant underuse of *take*-collocations in the NNS corpus, both in the number of types ($t=2.122$; $p=0.034$) and in the number of tokens ($t=2.067$; $p=0.039$).

	LANGUAGE	N	Types <i>n</i>	Mean	Stand. Dev.	Stand. error
TYPES	native	323	89	,78	1,181	,066
	non-native	251	45-**	,59	,887	,056

Table 5: Descriptive statistics and raw frequencies of collocation types based on NS and NNS corpora.

	LANGUAGE	N	Tokens <i>n</i>	Mean	Stand. Dev.	Stand. error
TOKENS	Native	323	318	,97	1,664	,093
	non-native	251	175-**	,73	1,159	,073

Table 6: Descriptive statistics and raw frequencies of collocation tokens based on NS and NNS corpora.

An analysis of the different types revealed that 23 types are common to both groups (table 7). This means that more than half of the types used by NNS are common to NS, that is, only 48.8 % are exclusive to NNS. From these, 6 types, that is 27%, have been classified as wrong combinations. As much as 74.1 % of the collocations produced by NS are exclusive to this group. However, if we calculate the type/token ratio for both native and non-native speakers,

similar results are obtained: NS: 0.28, NNS: 0.25. This seems to indicate that collocational patterns are not less diverse in NNS's than in NS's discourse.

Corpora	Native speakers only	Native speakers and learners	Learners Only	Total <i>n</i> of types
NS corpus	66 (74.1% of types in the NS corpus)	23 (25.8% of types in the NS corpus)		89
NNS corpus		23 (51.1% of types in the NNS corpus)	22 (48.8% of types in NNS corpus)	45

Table 7: Number and percentage of types exclusive to NS or NNS and types common to both.

3.5 Semantic analysis of collocations

A comparison of the frequency of the semantic categories defined for the study was carried out in order to see if there were differences between NS and NNS.

Semantic category	Total	Semantic category	NNS	Semantic category	NS
5	373	5	137	5	236
7	41	other uses	16	7	30
other uses	33	7	11	4	24
4	30	4	6	other uses	10
2	10	8	2	2	9
3	0	9	2	3	7
9	4	2	1	9	2
8	2	3	0	8	0
Total	493		175		318

Table 8: Semantic categories of collocations.

As we can see in table 8, the number of examples of category 5 (“delexical uses”) was by far the highest in both groups. Also, a similar ranking order of frequency of the different categories can be seen in both groups. No examples of category 3 (“phrasals”) were found in NNS data, whereas NS produced seven examples. Similarly, nine instances from the NS corpus were classified under category 2 (“eat or drink”), but only one from the NNS corpus.

3.6 Analysis of individual types of collocations

The individual analysis of the different types of collocations (cf. table 9) revealed that NNS underuse is general since, out of a total of 135 types, only two exhibited near-significant overuse by learners (*take place* and *take steps*).³ Eight types were significantly underused by learners and near-significant results were obtained for six more. These are shown in table 10.

³ Some of these types were common to both groups and some exclusive to one of them.

<i>SIGNIFICANT NNS UNDERUSE</i>	<i>t-value</i>	<i>p-value</i>	<i>NEAR SIGNIFICANT NNS UNDERUSE</i>	<i>t-value</i>	<i>p-value</i>
TIME/EFFORT ⁴	1.863	0.050	MEDICATION	1.897	0.059
action	2.909	0.004	bath	1.737	0.083
class	2.009	0.045	course	1.737	0.083
life	3.141	0.002	effect	1.737	0.083
option	2.009	0.045	precaution	1.737	0.083
position	2.009	0.045	responsibility	1.897	0.059
stance	2.009	0.045			
the easy way out	2.009	0.045			

Table 9: Individual analysis of the different types of collocations: NNS underuse.

<i>NEAR SIGNIFICANT NNS OVERUSE</i>	<i>T-value</i>	<i>p-value</i>
Steps	-1.683	0.093
Place	-1.715	0.087

Table 10: Individual analysis of the different types of collocations: NNS overuse.

Several reasons were considered to explain NNS underuse of collocations in general and of particular types. First, Granger (1998: 151) observed that “one could postulate that the learners’ underuse of *-ly* amplifiers is compensated for their overuse of *very*”. Similarly, since in our study *take* combinations are underused in general by NNS compared to NS, it is possible that in some cases the learner used a different verb instead of *take*, rendering, thus, a wrong or questionable collocation. Since, as we have observed, only *take* combinations are considered here, further analyses could clarify this point.

Another reason could be the underuse of the specific collocates. We extracted all the collocates occurrences belonging to the categories that exhibited significant or near-significant underuse in ICLE and selected those used in V+N combinations. This analysis was also meant to show whether other verbs instead of *take* had been incorrectly used for those collocates.

Finally, lack of congruency could also be a factor in learners’ underuse of particular types. Adopting Nesselhauf’s (2003: 236) definition, “only combinations that sounded natural in both languages if they were rendered word for word were regarded a congruent (considering, however, general syntactic rules of the two languages)”.

The analysis of individual examples did not yield conclusive results since no clear pattern was found. In some cases the collocate was not used at all by NNS; some of those were congruent and some non-congruent collocations.

Congruent

- precaution (“tomar precauciones”)
- the easy way out (“tomar el camino fácil”)
- stance (“tomar una postura”)

Non-congruent

- bath (“darse un baño”)
- responsibility (“asumir responsabilidad”)

In other cases, the collocate was found but always or much more often used in a different sense (i. e. one that required a different verb).

⁴ We grouped under this label the lemmas ‘time’ and ‘effort’ but also examples where different time or ‘effort’ expressions were used. Similarly, different types of medicines or drugs were grouped under MEDICINE, as well as the lemmas themselves.

Congruent

- position “tomar posiciones”
- TIME “me lleva XX hacer X”

Non-congruent

- life “quitar la vida”

Finally, there are a number of collocates, again both congruent and non congruent, which are found sometimes with *take*, sometimes with a different verb when *take* would be the target, and sometimes used in a different sense.

Non-congruent

- action: (“actuar”). 23 occurrences: with “take”. 1 example with a similar meaning with “carry out” instead of “take”.
- effect: (“surtir efecto”). 15 occurrences: 3 examples with a similar meaning with “produce” instead of “take”.
- option: (“optar” “inclinarse por una opción”). 6 occurrences. 1 example with a similar meaning with “choose” instead of “take”.
- effort: 20 occurrences: 1 with “take”. 4 occurrences with a similar meaning with “require” instead of “take”.

Congruent

- class: (“tomar clases”, but more frequent “ir a clase de”, which is incongruent). 9 occurrences: 1 example with a similar meaning with “join” instead of “take”.
- course: congruente (“tomar un curso”, but more frequent, “hacer un curso” which is incongruent). 10 occurrences: 1 with “take”. 2 with a similar meaning with “join” instead of “take”.
- drug: congruent “tomar una medicina / pastillas...” 8 occurrences: 3 examples occurrences with a similar with “have” instead of “take”.

On the one hand, although we could not find a single cause for the observed underuse of some types, we can conclude that it cannot be generally or exclusively ascribed to a lack of knowledge of the collocation or the use of a wrong verb, but to a more infrequent use of the collocate or to the use of the word in a sense that does not collocate with *take*.

On the other hand, two factors were considered to be determinant in the overuse of two types by NNS: frequency of the collocation in NS discourse and congruency. Thus, the combination *take place* is not congruent (“tener lugar”), but it is the most frequent type in the NS corpus. On the other hand, *take steps*, although not as frequent as the previous example, is nevertheless congruent (“tomar medidas”).

4. Conclusion

The intra-group analysis showed that most of the uses of *take* + noun combinations in both corpora are collocations (NS: 60.65%; NNS: 61.27%), which points to the predominance of phraseological uses of *take* by both NS and NNS as stated by Guilquin (2005).

The inter-group analysis of the different types of combinations revealed the underuse of collocations and free combinations on the part of the learners. This contrasts with their overuse of idioms. A detailed analysis of the types of idioms used by learners reveals that this overuse is due to the high frequency of a single type (i.e. *take into account*) which accounts for 31 out of 39, or 79.48% of occurrences of *take* idioms.

The individual analysis of the different types of collocations shows that the underuse of *take*-collocations is general, but only statistically significant in eight cases. Furthermore, although approximately half of the collocations used by NNS are also used by NS, the percentage of collocations which are exclusive to NS is rather high (74.1% of the NS corpus), which might point to the existence of an important gap in the learners’ active lexicon.

Several causes were considered to explain the underuse of particular examples of collocations, such as the use of another verb instead of *take*, underuse of the collocate, or lack of congruency with the L1 collocation. No conclusive results were achieved in some cases because further analyses beyond the scope of this paper would be needed, and, in some others, because no clear patterns were identified. However, the results indicate that the underuse of those types cannot be generally or exclusively ascribed to a lack of knowledge of the collocation or the use of a wrong verb, but to a more infrequent use of the collocate or to the use of the word in a sense that does not collocate with *take*.

As for the semantic analysis, category 5 (“delexical uses”) was by far the most frequently used by both groups. Similar ranking order of frequency of the different categories was observed in both groups, with small differences in categories 3 (“phrasals”) and 2 (“eat or drink”) which were more frequent in NS’s data.

This study is part of a larger research on the use of verb + noun collocations involving common verbs. Further research should be carried out on the verbs that learners use in contexts where a collocation with *take* is required. The study under description would be improved if it included a detailed description of the erroneous combinations produced by the learners and an analysis of the possible influence of the learners’ L1 on the production of both wrong and acceptable combinations.

References

- Altenberg B. and S. Granger. 2001. The Grammatical and Lexical Patterning of MAKE in Native and Non-native Student Writing. *Applied Linguistics* 22/2: 173-195.
- Bahns, J. 1993. Lexical Collocations: a Contrastive View. *English Language Teaching Journal* 47/1: 56-63.
- Bahns, J. and M. Eldaw. 1993. Should we Teach EFL Students Collocations? *System* 21/1: 101-114.
- Cambridge Advanced Learners’ Dictionary*
- De Cock, S. and S. Granger. 2004. High Frequency Words: the Bête Noire of Lexicographers and Learners Alike. A Close Look at the Verb “make” in Five Monolingual Learners Dictionaries of English. In G. Williams and S. Vessier (eds.), *Proceedings of the Eleventh EURALEX International Congress*. 233-243. Université de Bretagne-Sud: Lorient.
- Conklin, K. and N. Schmitt. 2008. Formulaic Sequences: Are they Processed More Quickly than Nonformulaic Language by Native and Nonnative Speakers? *Applied Linguistics* 29/1: 72-89.
- Cowie, A., ed. 1998. *Phraseology: Theory, Analysis, and Applications*. Oxford: Oxford University Press.
- Dayrell, C. 2007. A Quantitative Approach to Compare Collocational Patterns in Translated and Non-Translated Texts. *International Journal of Corpus Linguistics* 12/3: 375-414.
- Gabrys-Biskup, D. 1992. LI Influence on Learners' Renderings of English Collocations. A Polish/German Empirical Study. In P. J. Arnaud and H. Béjoint (eds.), *Vocabulary and Applied Linguistics*. 85-93. London: Macmillan.
- Granger, S. 1996. From CA to CIA and Back: An Integrated Approach to Computerized Bilingual and Learner Corpora. In K. Aijmer, B. Altenberg and M. Johansson (eds.), *Languages in Contrast. Text-based Cross-linguistic Studies*. 37-51. Lund: Lund University Press.

- Granger, S. 1998. Prefabricated Patterns in Advanced EFL Writing: Collocations and Formulae. In A. Cowie (ed.), *Phraseology. Theory, Analysis and Applications*. 145-160. Oxford: Clarendon Press.
- Granger, S., E. Dagneaux, and F. Meunier. 2002. *The International Corpus of Learner English. Handbook and CD-ROM*. Louvain-la-Neuve: Presses Universitaires de Louvain.
- Gouverneur, C. The Phraseological Patterns of High-Frequency Verbs in Advanced English for General Purposes. In F. Meunier and S. Granger (eds.), *Phraseology in Foreign Language Learning and Teaching*. 223-246. Amsterdam/Philadelphia: John Benjamins.
- Guilquin, G. 2005. To Take or not to Take Phraseology into Account. The Place of Multi-Word Expressions in Corpus Data and Experimental Data. In C. Cosme, C. Gouverneur, F. Meunier and M. Paquot (eds.), *Proceedings of the Phraseology 2005 Conference*. 165-168. Louvain-la-Neuve: Presses Universitaires de Louvain.
- Howarth, P. 1998. Phraseology and Second Language Proficiency. *Applied Linguistics* 19/1: 24-44.
- Hsu, J.T. and C. Chiu. 2008. Lexical Collocations and their Relation to Speaking Proficiency. *The Asian EFL Journal Quarterly* 10/1: 181-204.
- Källkvist, M. 1995. Lexical Errors Among Verbs: A Pilot Study of the Vocabulary of Advanced Swedish Learners of English. *Working Papers in English and Applied Linguistics* 2: 103-115.
- Kaszubski, P. 2000. Lexical profiling of English (learner) corpora: can we measure advancement levels? In B. Lewandowska-Tomaszczyk and J.P. Melia (eds.), *PALC'99: Practical Applications in Language Corpora. Papers from the International Conference at the University of Łódź, Poland, 15-18 April, 1999*. 249-286. Bern-Berlin-Bruxelles-Frankfurt/M.-New York-Oxford-Wien: Peter Lang.
- Keshavarz, M.H. and H. Salimi. 2007. Collocational Competence and Cloze Test Performance: a Study of Iranian EFL Learners. *International Journal of Applied Linguistics* 17/1: 81-92.
- Koya, T. 2004. Collocation Research Based on Corpora Collected from High School English Textbooks in Japan. In Y. Watanabe, I. Nagano, and A. Morita (eds.), *Collection of papers in honor of Professor Yoshiaki Shinoda*. 99-113. Tokyo: Nanundo.
- Leech, G. 1998. Preface. In S. Granger (ed.), *Learner English on Computer*. xiv-xx. London and New York: Addison Wesley Longman.
- Lorenz, G. R. 1999. *Adjective Intensification- Learners versus Native Speakers: A Corpus Study of Argumentative Writing*. Amsterdam: Rodopi.
- Nesselhauf, N. 2003. The Use of Collocations by Advanced Learners of English and Some Implications for Teaching. *Applied Linguistics* 24/2: 223-242.
- Nesselhauf, N. 2005. *Collocations in a Learner Corpus*. Amsterdam and Philadelphia: Benjamins.
- Orliac, B. 2008. Extracting Specialized Collocations Using Lexical Functions. In S. Granger and F. Meunier (eds.), *Phraseology. An Interdisciplinary Perspective*. 377-390. Amsterdam: John Benjamins.
- Oxford Collocations Dictionary for Students of English*
- Sinclair, J. 1991. *Corpus, Concordance, Collocation: Describing English language*. Oxford: Oxford University Press.
- Sung, J. 2003. *English Lexical Collocations and their Relation to Spoken Fluency of Adult Non-Native Speakers*. PhD dissertation. Pennsylvania University.
- van Roey, J. 1990. *French-English Contrastive Lexicology: An Introduction*. Louvain-la-Neuve: Peeters.

- Zhang, X. 1993. *English Collocations and their Effect on the Writing of Native and Non-native College Freshmen*. PhD dissertation. Indiana University.
- Zhang, W. and S. Chen. 2006. EFL Learners' Acquisition of English Adjective-Noun Collocations-A Quantitative Study. *Foreign Language Teaching and Research* 38/4: 251-258.