THE ACADEMIC WRITING OF UNIVERSITY ENGLISH LEARNERS: A CORPUS-BASED STUDY

A ESCRITA ACADÊMICA DE ESTUDANTES DE UNIVERSIDADE INGLESA: UM ESTUDO BASEADO EM CORPUS

Annallena de Souza Guedes
Universidade Federal de Minas Gerais

RESUMO: Este artigo discute e analisa a frequência e uso de verbos lexicais em textos argumentativos escritos por aprendizes brasileiros de Inglês da Universidade Federal de Minas Gerais. Utilizamos ferramentas da Linguística de Corpus (LC) para analisar os padrões gramaticais encontrados nos dados dos textos produzidos por aprendizes matriculados no IFA (Inglês para Fins Acadêmicos). Listamos os dez verbos mais freqüentes no corpus e, decidimos analisar o uso de *make* e *get* e, delimitamos os seus quinze colocados mais frequentes. Os resultados revelam que há dificuldades com o uso adequado desses verbos, especialmente no que se refere às colocações, bem como há uma tendência de interferência dos padrões de uso de sua língua nativa.

PALAVRAS-CHAVE: Escrita acadêmica; colocações; Make e Get; Linguística de Corpus; interferência.

ABSTRACT: This article discusses and analyses the frequency and use of high - frequency verbs in argumentative essays written by Brazilian English learners of Federal University of Minas Gerais. We used Corpus Linguistic (CL) tools to analyze the grammar patterns found on data of texts produced by the learners enrolled in English for Academic Purposes subject (Inglês para Fins Acadêmicos- IFA). We listed the ten most frequent verbs and we decided to analyze the use of *make* and *get* and outline their fifteen most frequent collocates. The results showed that there are some difficulties on the appropriate use of these verbs, especially when it turns to collocations, as well as there is a tendency of patterns use interference of their native language.

KEYWORDS: Academic writing; collocations; Make and Get; Corpus Linguistics.

Introduction

Corpora researches have shown that language can be used in a patterned way by its users, turning possible the association between use and context. As many studies have already pointed out, native speakers tend to use some patterns that non-native speakers do not. According to Cotos (2014), some studies have suggested that English language learners exhibit problems of frequency, register and phraseology when they produce texts.

Several recent studies lend support to this claim. Dutra & Silero (2010) carried out an investigation on grammar using learner corpora, in which they focused on the use of "for" in argumentative essays, by describing some aspects of learner interlanguage and the errors on writing, especially with the using of that preposition. Citing Halliday, the authors emphasize the importance of the empirical approach, in which language is

analyzed, firstly, by observing the frequency that motivate speakers / writers to choose certain lexical items rather than others in specific contexts of use.

In this way, it's important to highlight that this study is inserted in a descriptive enterprise, since our concern is not to provide rules for combining forms and consider deviations from the norm as errors (Biber, Conrad & Reppen, 1998), but provide an analysis about patterned use of grammar features by investigating the way learners use language in their constructions with two high-frequency verbs.

The availability of computers and a great amount of data have helped researchers to carry out studies about the patterned ways in which speakers use grammar features (Biber, Conrad & Reppen, 1998). This way, it's possible to investigate how languages are patterned in terms of their grammatical structures and categories, using for that the tools of Corpus Linguistics[1].

On this evidence, this study aims to investigate how the verbs *make* and *get* have been used in argumentative essays produced by Brazilian learners of English and the collocates that co-occur with them. This study analyzes Brazilian learner corpus CorIFA written by students at a federal public university in Brazil. They're all undergraduate and postgraduate students that have been taking an English for Academic Purposes classes, hereafter "IFA", a course offered to university students at undergraduate or post-graduate level. After the corpus analysis, we aim to compare the results with the ones obtained by Almeida (2014), in which she analysis the collocations produced by learners derived from the use of *get*, *make* and *take*.

Taking into account a study of Nesselhauf (2005), in which verbs constitute a major source of errors in free combinations and, considering that many studies have focused on learner writing, and more specifically, argumentative essay writing (Nesselhauf, 2005; Waibel, 2008), this study tries to analyze the high-frequency verbs *make* and *get* in CorIFA as well as their collocates, used in argumentative essays produced by the learners that have been already mentioned. Besides, it searches for giving a contribution to the English Language Teaching area, especially for teachers being aware of learners' specific writing difficulties and, accessing their students writing use them as a classroom resource (Cotos, 2014). Through this analysis, we hope to understand the impact of learners' native language interference on writing, since non-native speakers have a tendency of choosing

---

[1] There are two methodologies of investigation in Corpus Linguistics: Data-Driven approach and the Corpus-Based approach. For this study, we chose the former.

lexical items influenced by their native language (Gilquin, Granger & Paquot, 2007). Besides, we hope to raise some reflections on the teaching of English grammar, especially concerning to the need of making students aware of these grammar words, not as individual units, but as part of a construction within a phraseology. This concept will be discussed on the next section. This way, we aim at answering the following research questions: 1. Do the learners present some difficulties in the use of *make* and *get*; 2. If yes, which are them; 3. Which patterns of use are found in the corpus;

The article is structured as follows. In section 2, we highlight about the theoretical presuppositions that support this study, in which are included concepts on academic writing and grammar use. Section 3 presents the methodology used for such study followed by the obtained results and, in section 4, the final remarks, in which we demonstrate some of this study's implications in the teaching of English as an additional language.

1. Theoretical Presuppositions

In a study carried out by Yan (2006) with a high frequency verb, it's discussed how high-frequency words can express basic meanings and tend to be present in different semantic fields (Viberg, 1996). Besides, she states these words have equivalents in most languages as well as a high degree of polysemy, due to the universal nature of languages to create delexicalized or grammaticalized uses and the specific tendencies that result in meanings, collocations and idiomatic uses. The use of high-frequency verbs points to two arbitrary positions (Altenberg & Granger, 2001): an overuse of these verbs by learners or the underuse of them, and related to this former, Sinclair (1991, p.79) justify it by stating that "many learners avoid the common words as much as possible, and especially where they make up idiomatic phrases".

There have been many studies on academic writing *corpora* analysis whose main focus turns to the investigation of particular grammar points and the interference of learners' native language. In this direction, Almeida[2] (2014, p.77, our translation) believes that three factors may motivate the learners difficulties with phraseology. They are: "1. Interlingual factors that would envolve the negative transference of L1; 2. Intralingual

---

[2] In the original text: "1. Fatores interlinguais, que envolveriam a transferência negativa da L1; 2. Fatores intralinguais, referentes à falta de conhecimento na L2, principalmente a falta de conhecimento das restrições da língua e 3. Fatores relacionados ao ensino inadequado da L2".

Factors, referent to the lack of knowledge in the L2, mainly due to the lack of knowledge of language restrictions and 3. Factors related to the L2 inappropriate teaching". The author pointed out L1 transference as one of the main motivations for the collocational errors on learners' texts.

Many studies on learner corpora have been carried out worrying about phraseology. They show that understanding words as non-isolated units of sense could potentialize the difficulties learners have when it turns to writing, more strongly, when they write academic texts, since words have to be considered within a phraseology. In this direction, Hunston (2009) raises the concept of phraseology as a determinant factor to meaning-making, as well as she discusses about the *idiom principle* postulated by Sinclair (1991). According to her, there is no distinction between *pattern* and *meaning* and, between lexis and *grammar*. The phraseology consists not on a set of joined phrases, as words do not occur randomly, but a set of words that, joined together, allows meaning construction.

Learner corpora gained popularity with Granger's (1998) work and can be defined as:

> "(...) electronic collections of authentic texts produced by L2 learners, can help to reveal those difficulties and to understand the differences between learner production and the features that characterize native-like language use" (COTOS, 2014, p.203).

Learner corpora research has helped the linguistics field by identifying features of language phenomena associate to language produced by non-native English speakers, since their vocabulary repertoire is short when compared to native speakers (Gilquin, Granger & Paquot, 2007). Thus, dealing with learner corpora therefore, allows the researcher to identify learners use and performance in the foreign language. (Shepherd, 2009). Studies on grammar patterns investigations focus on "how languages are structured and how minds are working as they produce and process language" (Biber, Conrad & Reppen, 1998, p.56).

There's a strong debate among researchers on acquisition of a non-native language about teaching or not teaching grammar (Meunier, 2002). Teachers must help learners to realize about phraseology and its importance to language learning and proficiency (Granger & Meunier, 2008).

2. Methodology

Data analyzed come from CorIFA, composed of undergraduate and postgraduate students' argumentative essays. They are ongoing *corpora*, which have been compiled since 2012. At the time of the data analysis the former was composed by 148.843 words. The prompts for the essays covered a variety of topics and they were all proposed by the teachers and required for groups on the same level.

Data were analyzed according to some procedures. Firstly, we selected the *corpus* to verify how the investigated grammar patterns occur on them. After that, we devoted to the verbs identification as well as their first three collocates in the left and in the right that make part of their phraseology.

In this work, data handling is carried out in a semi-automatic way, through the freeware concordance *AntConc*[3], since we could check the words frequency and collocates, through analyzing the concordance lines in which these grammar constructions occurred. We used the Concordance tool to generate the concordance lines with the verbs, as well as to collect all the collocates of *get* and *make* in the corpus.

2.1 Data collection and analysis

In the CorIFA, we found the following ten most frequent verbs: *have, be, can, do, like, live, make, would, use* and *could*. As a criterion for the analysis, we decided to analyze only two of the ones presented in the chart: *make* and *get*.

MOST FREQUENT LEXICAL VERBS
Have
Live
Make
Use
Do
Learn
Know
Need
Get
Work

Table 1. The most frequent verbs in the CorIFA

---

[3]The software, created by Laurence Anthony, is available for free from http://www.antlab.sci.waseda.ac.jp/software.html.

It's noticeable that all these verbs are high-frequency ones in English, especially in written registers of school contexts. In fact, most of them are considered high-frequency ones in a list of fifteen according to some corpus-based studies: *have, go, take, do, say, look, know, see, give, think, come, find, get, make* and *use*. Six of these verbs were also found as the ten most frequent ones in CorIFA. We will analyze just two of them, showing their position and patterns of use in the concordance lines as well as their collocates.

The concordance shows any word in its cotext (Sinclair, 1991) of use referring to the linguistic environment in which the word is situated and, the words units that are used on its left and right. The word in the center of the lines, printed in blue, correspond to the node and the items around it are called collocates. In our study, we analyzed just the items in the right position from the node.

In the next section, we analyze the uses of *make* and *get* found through their respective concordance lines in the corpus.

3.1.1 Uses of *make*

The verb *make* has many different meanings and uses (Altenberg & Granger, 2001, p.177). The most distinctive ones are the delexical and the causative uses. In this investigation, we will concentrate on the analysis of both since they're the most representative ones in the corpus.

In a study carried out by Almeida (2014) to identify and analyze collocations in learner corpora and compare her findings with COCA[4] and BR-ICLE[5], she realized that *make* is largely used in the sense of perform, which demonstrates that Brazilian learners, even in academic discourse texts, do not use academic verbs, preferring this way, the high - frequency ones.

Similar results can be seen in CorIFA. In the concordance lines below these findings are registered:

---

[4] Corpus of Contemporary American English. Available from http://corpus.byu.edu/coca/.
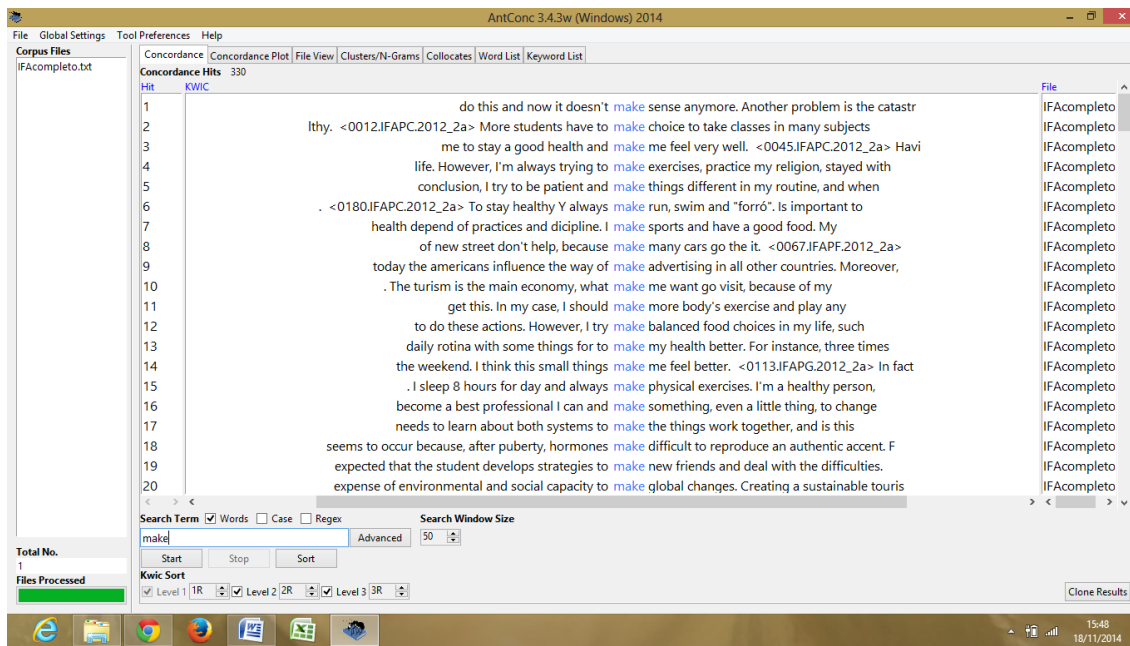[5] Brazilian Sub-Corpus of the International Corpus of Learner English - ICLE.

Figure 1 –
Screen of concordance lines for the verb *make*
Source: AntConc, ANTHONY, 2011.

She also found out that Brazilian learners often use the causative pattern of *make: make life easier, make people sick, make us believe*, etc. The table below brings out the use of *make* as a causative verb in the studied corpus:

Causative uses of *make*		
Use of adjective structures	**Use of verbal structures**	**Use of nominal structures**
make things different	*make me feel*	*make the things*
make my health better	*make me want*	
make me proud and patriot	*make me grow*	

Table 2. Causative uses of *make*

Explaining the frequent use of causative *make*, Almeida (2014) believes that it can be due to interlinguistic and intralinguistic aspects as Altenberg & Granger (2001) have pointed out. In relation to the interlinguistic aspects, we found the pattern *make+noun+adjective* as in *make things different.* In this case, the adjective plays an important role in meaning construction as well as in *make me proud* (make someone proud of something) and in the other examples in which there are adjective structures. This is also a construction found in these learners' native language.

Another pattern in the chart above is related to *make +pronoun+verb*, indicating the verb as an important category in the construction, with a meaning of *make someone do something* as in *make me want* and *make me grow.*

7

The verb make is also used with a number of different nouns, indicating noun expressing speech actions, which represent its delexical structures (Sinclair, 1990). On this evidence, we found the following uses in the studied corpus:

Delexical uses of *make*		
make a appointment	*make advertising*	*make the things*
make a call	*make global changes*	*make friends*
make choice	*make a dinner*	*make sports*
make exercises	*make a barbecue*	

Table 3. Delexical uses of *make*

As it can be gleaned from it, there are some uses that do not present collocational misuse, such as *make a appointment* (even though the use of the article is not right), *make a call* and *make a choice*. In these cases, the verb *make* combines with direct objects and build phrases whose meaning widely depends on the meaning of the noun.

On the other hand, in the construction *make exercises*, the learner used it as a meaning of "practice physical activities". Consulting the reference corpus COCA (Corpus of Contemporary American English), we found only two occurrences of *make exercises*, while the construction *do exercises* has 48 occurrences, which corroborates to the idea that the former can be the right pattern of such construction in the context of practice physical activities. The same happens to the other uses: *make a barbecue* (1 occurrence), *make advertising* (4 occurrences), *make the things* (13 occurrences), *make sports* (18 occurrences) and *make a dinner* (19 occurrences). The other uses *make a call* (218 occurrences), *make a choice* (407 occurrences) and *make friends* (643 occurrences) presented a high frequency use pointing out to the pattern of the language use.

These results can be an indication of "a combined effect of interlingual and intralingual forces: languages have a dominant pattern with equivalent high-frequency verbs" (Altenberg & Granger, 2001, p.182).

The chart below shows the fifteen most frequent collocates of the verb *make.* There are 35 collocate types and a total number of 494 collocate occurrences in the corpus, considering words that co-occur three places to the right and a minimum of 5 collocate frequency.

Rank	Frequency (R)	Collocates
1	56	*the*
2	49	*a*
3	36	*and*
4	23	*our*
5	21	*easier*
6	19	*life*
7	18	*to*
8	17	*better*
9	16	*friends*
10	15	*more*
11	14	*world*
12	13	*in*
13	12	*for*
14	11	*this*
15	10	*good*

Table 4. Frequencies of the fifteen most frequent right collocates of *make*

Such frequency identifies the most common combinations with the word *make* in the corpus, called collocates and they give a semantic profile (Hunston, 2002) of the verb involved. The figure, in decreasing order of frequency, points to clear frequency of prepositions and pronouns as important words to the behavior of *make: our, to, in* and *for.* It's also important to note that the frequencies are not so variable. So, the most frequent collocates are definite article (*the*), the indefinite article (*a*) and the conjunction (*and*). Besides, the verb collocates with words that come from a positive semantic environment, such as *easier, life, better, friends, more* and *good.*

Considering one of these collocates, it's also found the pattern *make + article + adjective + noun* occurring in most part of the concordance lines below:
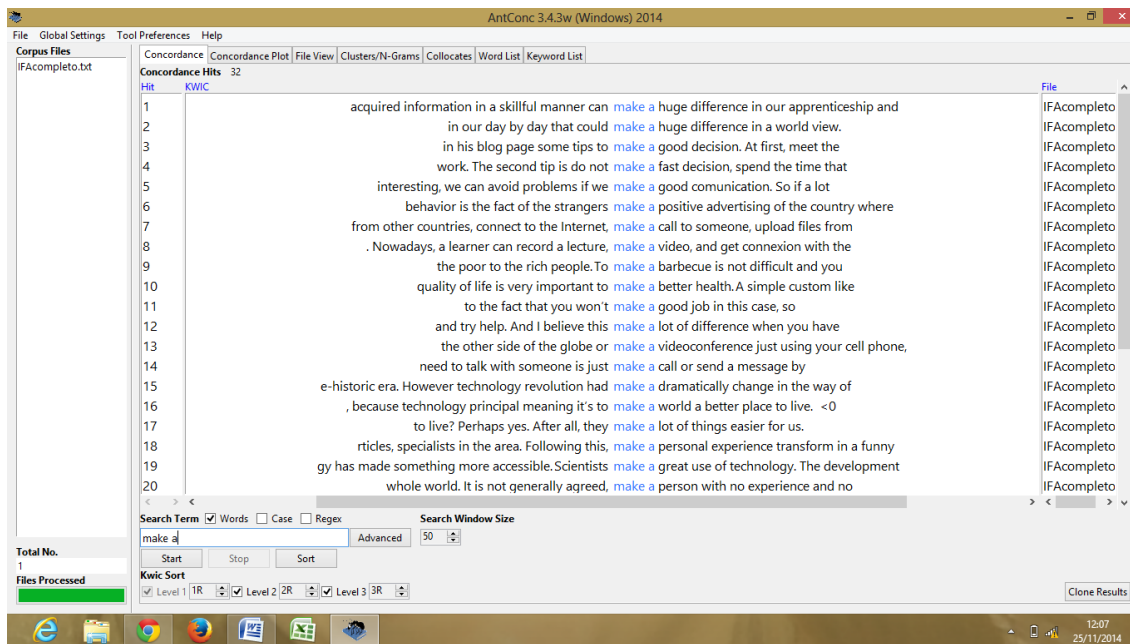
Figure 2 –
Screen of concordance lines for the verb *make*
Source: AntConc, ANTHONY, 2011.

The constructions *make a huge difference*, *make a good job* and *make a great use* are examples of this recurrent pattern in the corpus. Acknowledging this fact, we observe that learners transfer their native language pattern when they produce texts in an additional language, since in Portuguese these constructions are very similar and would correspond to "to cause something to happen", "perform a good task" and "use something greatly". The construction *make a decision* cannot be replaced by the synonymous *do* or *produce*, since word combinations have arbitrary restriction on the commutability of their words.

3.1.2 Uses of *get*

As well as make, the verb *get*, due to its polysemous and phraseologic nature, has many different uses. In the picture below, it's shown the concordance lines of the verb *get*:
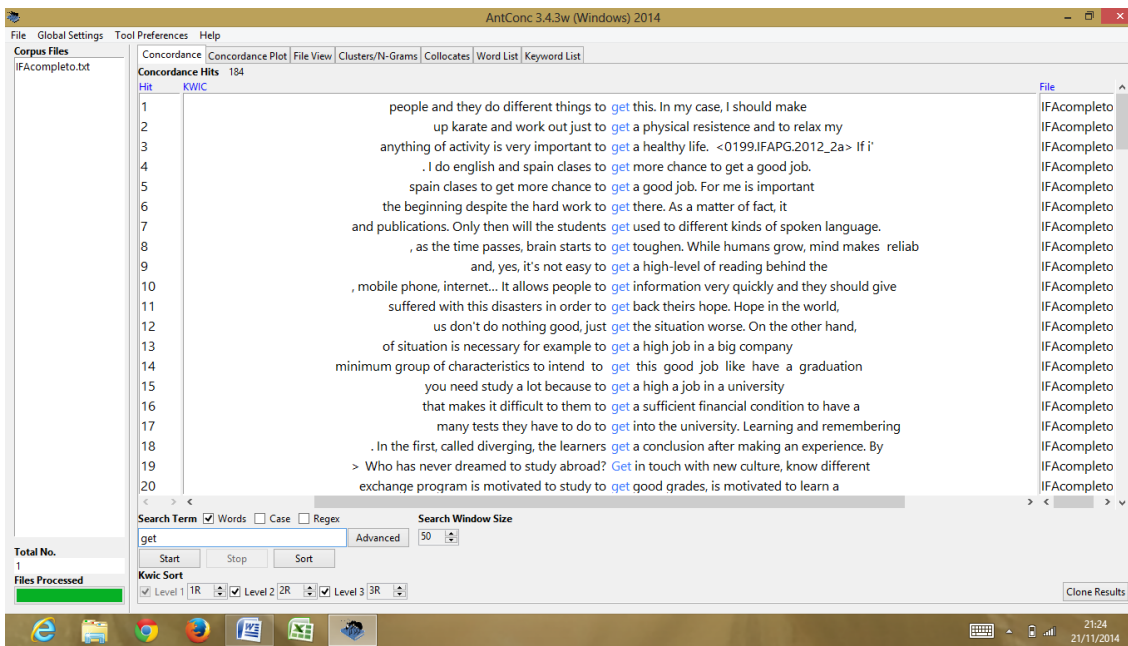
Figure 3 –
Screen of concordance lines for the verb *get*
Source: AntConc, ANTHONY, 2011.

As we could see in the concordance lines, the verb "get", mostly used followed by an article occur 8 times (only in the screen showed above), which indicates a pattern (*get + article + adjective + noun*) as in the following examples from the corpus:

GET + ARTICLE + ADJECTIVE+NOUN
get a physical resistance
get a healthy life
get a good job
get a high job
get a sufficient financial condition

Table 5. Examples with *get + a* taken from the concordance lines

They are also typical and common expressions, often taught for first levels groups. This fact corroborates with a tendency learners have of choosing and using lexical items which they're closely familiar with.

In order to better visualize the combination of words and the pattern, we got the following result:
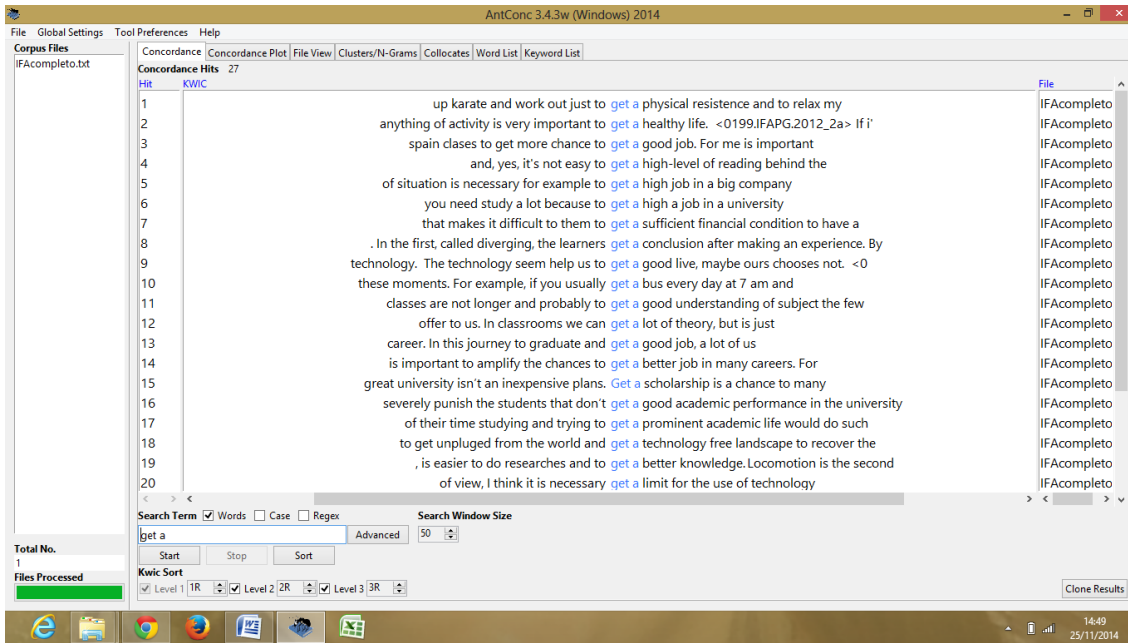
Figure 4. Concordance lines of *get* identifying the patterns verb+article+adjective + noun and verb+article+noun.

The pattern is the most frequent one in the corpus with 27 occurrences. In total, there are 20 collocate types and a total of 236 collocate tokens. These combinations were analyzed one by one, in order to identify how Brazilian learners use *get* in their written productions. From all these combinations, few of them were idiomatic expressions. This way, considering words that co-occur three places to the right and a minimum of 5 as a collocate frequency, these results were found:

Rank	Frequency (R)	Collocates
1	32	a
2	29	to
3	29	the
4	23	in
5	15	and
6	11	better
7	11	good
8	11	more
9	9	information
10	9	of
11	7	on
12	6	grades
13	6	job
14	6	some
15	6	this

Table 6. Frequencies of the fifteen most frequent right collocates of *get*

The frequency identifies the most common combinations with the word *get* in the corpus. The figure, in decreasing order of frequency, points to clear frequency of articles, prepositions and adjectives as important words to the verb behavior*: a, to, the,* and *good.* The nouns *information*, *grades* and *job* are related to the constructions "get information", "get grades" and "get a job", which, according to Oxford Advanced Learners' Dictionary mean "receive something", "to achieve or be given a particular grade" and "to achieve an opportunity", respectively.
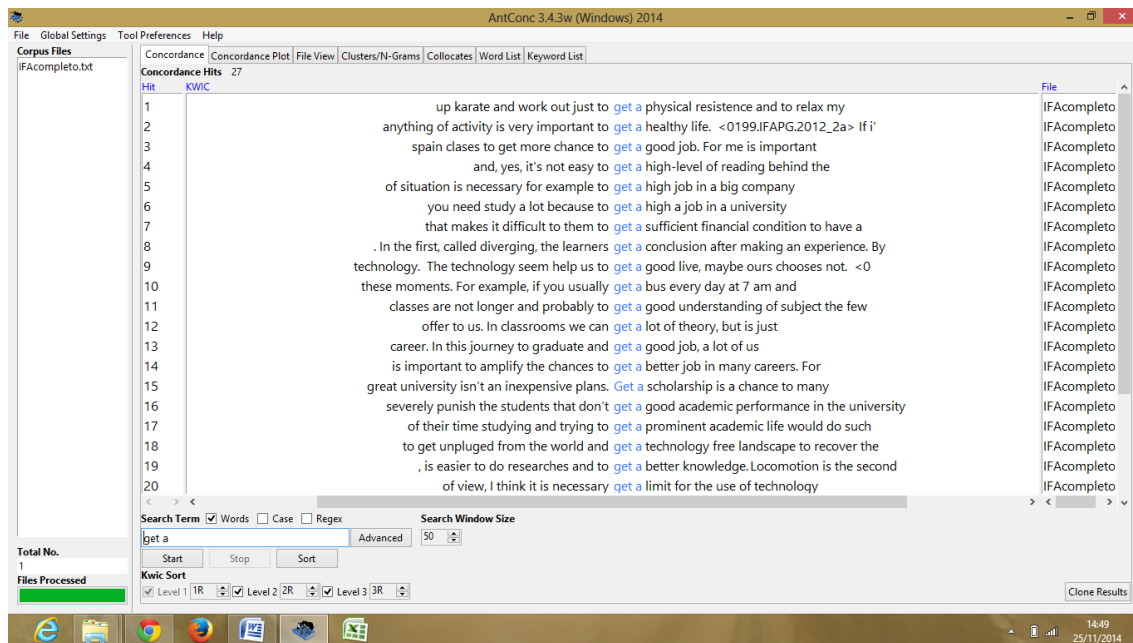

Figure 5.
Concordance lines of the pattern *get + a*

Some misuses are found in these concordance lines. The constructions *get a lot of theory, get a better knowledge* and *get a high job* are examples of learners' native language interference since they tried to refer to "learn some theoretical pressupositions", "learn better" and "get a good job" respectively. We verified theses uses found in CorIFA in the reference corpus COCA and realized no occurrences of *get a lot of theory* and *get a high job* and, at last, there is just one token of *get a better knowledge.* When observing the construction *get a good job* in COCA, there were 84 tokens, which corroborates to the idea that this is a more patterned use among the native speakers.

3. Results

The analysis of the high-frequency verbs *make* and *get* confirms findings of similar studies investigating collocational use with them, like Almeida's (2014) investigation. Learners tend to overuse some expressions and words they are familiar with since other expressions which carry the same function are unknown for most of them.

Besides, as it had been already pointed out by Almeida (2014), learners have some difficulties on using high - frequency verbs, such as do, get, give, have, make, take, aming others. This way, the results corroborate to reinforce what Nesselhauf (2005) has stated about the difficulties learners have in combining words in written productions.

4. Final remarks

Pinto (2012) believes that learner corpora research highlights the importance of changing paradigms in the process of teaching additional languages since students' written productions seem to indicate an innapropriate use of lexicogrammar patterns.

The use people do with the words is sistemically patterned and this corroborates to what Biber, Conrad & Reppen (1998) have established. It's important, this way, to study the meaning and use of words like verbs, in order to investigate why some of them occur so commonly, reinforcing the relevance of considering the different forms of the word collectively.

Furthermore, corpus-based studies may enhance the way textbooks, for example, present and teach grammatical constructions, by emphasizing the most common ones in the target register, as well as presenting the communicative purposes of them.

Learner corpora can provide a new type of data which can inform thinking both in SLA (Second Language Acquisition) research, since it searches to understand the mechanisms of foreign/second language acquisition, and in FLT (Foreign Language Teaching) research, the aim of which is to contribute to the enhancement of teaching and learning of foreign/second languages outcomes (Granger, 2002).

As the analysis was limited in scope its findings point to several further research directions, especially in the need of investigating grammatical accuracy with the use of determiners, prepositions and collocations.

References

Almeida, V.C. (2014). *Investigando colocações em um corpus de aprendiz*. Tese de doutorado, Universidade Federal de Minas Gerais, MG, Brasil.

Altenberg, B.& Granger,S. (2001). The Grammatical and Lexical Patterning of MAKE in Native and Non-Native Student Writing. *Applied Linguistics* 22/2: 173-195

Anthony, L. (2014) AntConc   (Version 3.2.4) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.antlab.sci.waseda.ac.jp/.

Cotos, E. (2014). Enhancing writing pedagogy with learner corpus data . *ReCALL*, 26, pp 202-224.

Dutra, D.P & Silero, R.P. (2010). Descobertas linguísticas para pesquisadores e aprendizes: a linguística de corpus e o ensino de gramática. *Revista Brasileira de Linguística Aplicada*, 10, 4, Belo Horizonte.

Biber,D.; Conrad, S. & Reppen, R. (1998). *Corpus linguistics. Investigating language structure and use.* Cambridge: Cambridge University Press.

Gilquin,G.; Granger,S.& Paquot, M. (2007). Learner corpora: the missing link in EAP pedagogy. *Journal of English for Academic Purposes*, 6, 4, pp 319-335.

Granger, S. (2002). *A bird's eye view of learner corpus research*. In Granger, S; Hung, J. & Petch-Tyson, S. (eds.). Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching. Amsterdam: John Benjamins, PP 3–33.

Granger,S.; Meunier,F. (2008). Phraseology in Foreign Language Learning and Teaching. Amsterdam: John Benjamins Publishing Company.

Hunston, S. (2002). *Corpora in Applied Linguistics.* Cambridge: Cambridge University Press.

Hunston, S. (2009). *Corpora na language teaching: Issues of language description*. In Corpora in Applied Linguistics. Cambridge: Cambridge University Press.

Meunier, F. (2002). The Pedagogical Value of Native and Learner Corpora in EFL Grammar Teaching. In: Granger/Hung/Petch-Tyson 2002, 119141.

Nesselhauf, N. (2004). *Learner corpora and their potential for language teaching*. In Sinclair, J. (ed.), How to use corpora in language teaching. Amsterdam: John Benjamins Publishing Company, pp 125-152.

Nesselhauf, N. (2005). *Collocations in a Learner Corpus*. Amsterdam: John Benjamins Publishing Company.

Pinto, M.V. (2012). *O uso de things, thing, anything, something e everything em corpus de aprendiz.* In Shepherd,T.M.G; Sardinha,T.B; Pinto, M.V (orgs). Caminhos da Linguística de Corpus. São Paulo: Mercado das Letras.

Shepherd,T. M.G. Corpora de aprendiz de língua estrangeira: um estudo contrastivo de n-gramas.Veredas on-line – Linguística de Corpus e Computacional - 2/2009, pp. 100-116.

Sinclair, J. (ed.), 1990. *Collins Cobuild English Grammar*. Glasgow: Harper Collins Publishers.

Sinclair, J. (1991). *Corpus, Concordance, Collocation.* Oxford: OUP.

Viberg, A. (1996). *Cross−linguistic lexicology. The case of English go and Swedishga*. In K. Aijmer, B. Altenberg & M . Johansson（eds）. Languages in Contrast. Lund : Lund University Press.

Waibel, B. (2008). *Phrasal verbs. German and Italian learners of English compared*. Saarbrücken: VDM.

Yan, Q. (2006). A corpus-based analysis of the verb "do" used by Chinese Learners of English. *CELEA Journal*, 29,6.