English collocations extracted from a corpus of university learners and its contribution to a language teaching pedagogy

Adriane Orenha-Ottaiano

Departamento de Letras Modernas, Instituto de Biociências Letras e Ciências Exatas, Universidade Estadual Paulista "Júlio de Mesquita Filho", Rua Cristóvão Colombo, 2265, 15054-000, São Jose do Rio Preto, São Paulo, Brazil. E-mail: adrianeorenha@gmail.com

ABSTRACT. Taking into consideration the relevance of foreign language teaching and the learning of collocations (ALTENBERG; EEG-OLOFSSON, 1990; FONTENELLE, 1994; MEUNIER; GRANGER, 2008), this paper aims at showing results of an investigation on whether the teaching of collocations should be implicit or explicit to the Brazilian university students. Furthermore, the research has the purpose of presenting some collocational aspects from a *corpus* of the written language learners made up of intermediate, upper intermediate and advanced university students' argumentative essays at a public university in Brazil. With the help of *WordSmith Tools* (SCOTT, 2007), it was possible to raise students' most frequent collocational choices and patterns, the most/least used type of collocations, the influence of the mother tongue on their choices, among other aspects. With the purpose of motivating and involving students in classroom research, it was also introduced *The Corpus of Contemporary American English* (COCA), created by Mark Davies. By doing so, students could compare their collocational choices with the patterns found in the online corpus, extract more collocational patterns and, consequently, be aware of the potential of corpora for the foreign learning process, specifically for raising language awareness, with focus on prefabricated chunks.

Keywords: collocational pattern, online corpora, phraseology, foreign language.

Colocações em inglês extraídas de um *corpus* de aprendizes universitários e sua contribuição para uma pedagogia no ensino de línguas

RESUMO. Considerando a relevância das colocações para o ensino-aprendizagem de língua estrangeira (ALTENBERG; EEG-OLOFSSON, 1990; FONTENELLE, 1994; MEUNIER; GRANGER, 2008), este artigo tem como propósito mostrar os resultados de uma investigação a respeito do ensino implícito ou explícito de colocações para alunos universitários brasileiros. Além disso, a pesquisa visa apresentar alguns aspectos colocacionais, levantados a partir de um *corpus* de aprendizes de língua escrita, composto por textos argumentativos de alunos de nível intermediário, intermediário avançado e avançado, de uma universidade pública, no Brasil. Com o auxílio do programa computacional *WordSmith Tools* (SCOTT, 2007), foi possível levantar as escolhas e os padrões colocacionais mais frequentemente empregados pelos alunos, os tipos de colocações mais/menos frequentemente empregadas, a influência da língua materna nessas escolhas, entre outros aspectos. Com o objetivo de motivar e envolver os alunos em pesquisa em sala de aula, foi apresentado aos alunos *The Corpus of Contemporary American English* (COCA), criado por Mark Davies. Ao utilizá-lo, os alunos puderam comparar suas escolhas colocacionais com os padrões encontrados no *corpus*, extrair outros padrões colocacionais e, consequentemente, conscientizar-se do potencial de corpora no processo do aprendizado de línguas estrangeiras, com foco em blocos pré-fabricados.

Palavras-chave: padrão colocacional, corpora on-line, fraseologia, língua estrangeira.

Introduction

Collocations, under the scope of conventionality and being one among various types of phraseologisms, play an important role in the use of a language. Many researchers (ALTENBERG; EEG-OLOFSSON, 1990; FONTENELLE, 1994; MEUNIER; GRANGER, 2008) claim that learners involved in spontaneous interactions constantly need expressions easily retrieved from their mental lexicon, besides vast repertoires of preferable ways of saying things. The larger the stock of lexical combinations a learner has at their disposal, as well as a series of expressions related to them, the smaller their effort to code and decode.

Due to that, as stated by Altenberg and Eeg-Olofsson (1990, p. 2, author's underline), these prefabs or lexical combinations act "[…] as a kind of 'autopilot' which the speaker can switch on to gain time for the

creative and social aspects of the speech process". Taking this context into account, the learning of collocations and other prefabricated chunks becomes crucial to learners whose aim is to produce more fluent and conventional speech. Furthermore, researchers (GRANGER, 1998; ORENHA-OTTAIANO, 2009; SINCLAIR, 1991) claim that the teaching of collocations can be enhanced by the use of corpora in foreign language classroom.

Therefore, once observed the relevance of collocations in foreign language learning and teaching, this paper aims, as one of its purposes, to shed light on collocational patterns, by discussing a research question: should the teaching of collocations be implicit (incidental), that is to say, should they be learned in a more or less automatic way like other lexical items, or should they be taught in a more explicit way (intentional)?

According to partial results from our research carried out in a public University in Brazil, students did not have good collocational test results when collocations were not taught explicitly, confirming other researches which also support the explicit teaching of collocations (BAHNS; ELDAW 1993; CHANNEL, 1981; CONZETT, 2000; FONTENELLE, 1994; HILL, 2000; MARTON, 1977). It is also worth noting this is the first phraseological investigation undertaken among Brazilian university learners of English whose learning, knowledge of, and fluency in the referred language is crucial for their success in their career, as some of them will hold a B.A. degree in Translation and others will have a B.A. in the English Language Teaching. This situational context may have a significant effect on the essays that make up the studied corpus and that fact also makes it to be the first corpus in Brazil that is entirely composed of texts written by future professionals of the language. To our knowledge, there are two other learners' corpora being compiled or already completed here in Brazil, whose subjects are different from the ones involved in this investigation. The first corpus is the *Br-Icle*, compiled at the Catholic University of São Paulo under the coordination of Professor Antonio Berber Sardinha, and composed of essays written by undergraduates from different majors. The other one is *The USP Multilingual Learner Corpus*, made up of English, French, Italian, German and Spanish texts produced by their undergraduate students, students, teachers and other employees who take language classes in *campus* courses at USP, with different language, educational and professional background.

Besides that goal, this research has focused on the building of a written learner corpus, composed of argumentative essays based on themes discussed during their writing classes. With the help of the computing program WordSmith Tools (SCOTT, 2007), version 5.0, it was possible to raise students' most frequently used collocations and have some partial results on: 1) the students' collocational choices and patterns; 2) the influence of the mother tongue on these choices; and 3) the most/least used type of collocations employed by the Brazilian students: verbal, nominal, adjectival or adverbial collocations.

It will also be discussed some advantages of using online corpora in the foreign language classroom as a motivating and awareness factor.

Literature review

Computer learners' corpus

According to Granger et al. (2002, p. 7), computer learners' corpora are "[...] electronic collections of authentic FL/SL textual data assembled according to explicit design criteria for a particular SLA/FLT purpose". One of the most significant advantages of learner corpora is the fact that one can have a record of the learners' production which may enable researchers to report what learners can actually produce, for instance, in terms of lexicogrammar patterns and phraseological aspects.

Among the learner corpora that are current available, the most well-known are: *The International Corpus of Learners of English* (ICLE), under the supervision of Professor Sylviane Granger (1993), *The Longman Learners' Corpus*, by Longman, *The Cambridge Learners' Corpus*, under the responsibility of the Cambridge University Press and Cambridge ESOL, among many others, listed on the site of Université Catholique de Louvain[1].

Regarding the applications of learner corpora to the foreign language teaching, researchers (AIJMER, 2009; FRANKENBERG-GARCIA et al., 2011; GRANGER, 1998; GRANGER et al., 2002; NESSELHAUF, 2005; TONO, 1999) highlight various advantages of using them. They can have access not only to learners' errors, but also to learners' interlanguage (interlanguage description). They can also exploit them for production of foreign language material and the compilation of dictionaries. Besides that, according to Pravec (2002), the learners' corpora enable the investigation of 'foreign soundingness' in non-native essays by

---

[1]The Université Catholique de Louvain site is available from: <http://www.uclouvain. be/en-cecl-lcWorld.html>.

analyzing which grammatical, linguistic, lexical, or pragmatic structures may be overused or underused with regard to the target language norm.

Meunier et al. (2010) show how the learners' corpora research has enormously contributed to the learning and teaching of foreign languages and discuss the numerous linguistic and pedagogical benefits corpora may bring to the area. De Cock et al. (1998), Granger (1998), and Granger et al. (2002) deal with the influence of the mother tongue on learners' output of multiword sequences. In addition, De Cock et al. (1998) states that exploiting a learners' corpus allows us to find out which areas a learner from a specific country needs help for developing their writing skills.

Collocations

Collocations are one of the various types of phraseologisms belonging to the scope of conventionality, as it is shown in the Figure 1:
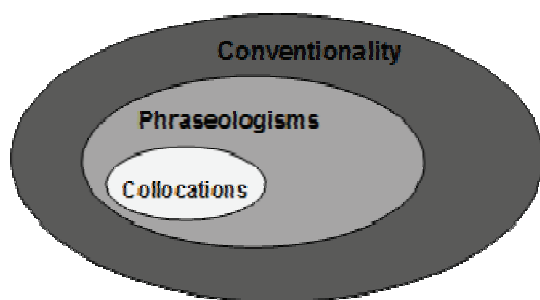

Figure 1. Conventionality, phraseologisms and collocations.
Fonte: Orenha-Ottaiano (2004, p. 13).

Being in the realms of conventionality, collocations are not considered to be a problem regarding comprehension, but production. That means it is, in most cases, perfectly possible to understand them. For example, a learner may easily understand the collocation 'to place an order' and 'to pay a compliment', as long as they know the meaning of 'order e compliment'. However, an elementary or pre-intermediate learner would have difficulties in producing them, and very often influenced by their mother tongue, would end up making collocational errors: Brazilian learners would tend to use either the verb 'to do' or 'to make' and produce collocations such as 'make an order' or 'make a compliment'. That would probably be understood by a native speaker, but such constructions would sound fairly strange.

Hausmann (1985) has greatly contributed to the description and definition of collocations. According to the author, collocations are 'semi-finished products of a language', and their most important aspect is their "[…] status of mental disponibility as a whole, not as a creation produced ad hoc by a speaker" (HAUSMANN, 1984 apud HEID et al., 1991, p. 15). Hausmann claims that speakers of a language simply reuse the 'semi-finished products of a language' when they use collocations.

With regard to the elements of a collocation, Hausmann (1984) points out two: a basis and a collocate, each with different semantic status. Actually, there is a hierarchy between these two elements, as one of them determines (the basis) and the other is determined (the collocate). Simply speaking, the basis is what we already know and the collocate is the element we are looking for. The basis is an independent element, semantically autonomous, that determines which lexical patterns can combine with it. On the other hand, the collocate works as a modifier, it is semantically interpretable within a collocation and it is chosen by a certain basis to form a collocation (HEID et al., 1991).

Concerning the taxonomy of collocations, Hausmann (1985) suggests a classification which was expanded by Orenha-Ottaiano (2009), whose examples were taken from a corpus of business English:

Verb-like – with five basic forms:
- Verb ~collocate~ + Noun ~basis~: *acquire shares*
- Noun ~basis~ + Verb ~collocate~: *investments dropped*
- Verb ~collocate~ + Preposition + Noun ~basis~: *dispose of shares*
- Verb ~collocate~ + Adverbial Particle + Noun ~basis~: *set up a business*
- Verb ~collocate~ + Adjective ~basis~: *grow strong*

Name-like – with two basic forms:
- Noun ~basis~ + Noun ~collocate~: *share subscription*
- Noun ~collocate~ + Preposition + Noun ~basis~: *holder of shares*

Adjective-like – with one form:
- Adjective ~collocate~ + Noun ~basis~: *bearer shares*

Adverb-like – with three basic forms:
- Adverb ~collocate~ + Adjective ~basis~: *fully eligible*
- Verb ~basis~ + Adverb ~collocate~: *drop dramatically*
- Adverb ~collocate~ + Verb ~basis~: *fully paid*; *duly appointed*

For the current paper, it was possible to present verbal, nominal and adverbial collocations extracted from the University Learners' Corpus and COCA Academic section.

The subjects and the compilation of the University Learners' Corpus

The subjects of this research are university students from the 1st and 2nd year of a B.A. in an English Language Teaching Course and the 3rd and 4th year of a B.A. in a Translation Course, both courses from a public University in Brazil. In the near future, it is intended to include the 1st and 2nd year of the Translation Course's and the 1st and 2nd year of the B.A. in English Language Teaching's essays to be stored in the corpus. For the first part of the paper, the focus will be on the 2nd year of the B.A. in the English Language Teaching Course's and the 3rd and 4th year of the Translation Course's collocational production, as there's not enough data from the 1st year students yet. For the second part of our analysis, on the collocational aspects observed in the University Learners' Corpus, it was selected as a sub-corpus of the 3rd and 4th year students from the Translation Course because it is mainly made up of argumentative essays. It is important to mention that the classes, from both courses, have a very heterogeneous level of proficiency. The students come to the university with different English background knowledge, as some of them have already studied English in language schools. Thus, it is possible to have intermediate, upper intermediate and advanced students all together in the second year of the B.A. in the English Language Teaching or the Translation Course, for example.

Concerning the compilation of the University Learners' Corpus, it is composed of argumentative essays written by the above mentioned university students as Writing Class assignments, following some procedures proposed by Granger (1993). Similar to the ICLE (GRANGER, 1993, 1998), the students also have to write on previously selected themes. As the corpus is still under compilation and revision, especially the sub-corpus from the 1st and 2nd year of the B.A. in the English Language Teaching Course, it was chosen the sub-corpus of the 3rd and 4th year students to be analyzed in the scope of this paper. Here are some of the themes already chosen and compiled for the referred group (Table 1).

Table 1. Topics for the essay writings.

3rd year	4th year
Happiness	Genetically-modified food
Stereotypes	Same-sex marriage
Global Warming	Addiction/Drugs
Culture Shock	Euthanasia
Religion	Capital Punishment
Censorship	Advertisement

In what refers to the number of words of each essay, it was decided on 500 to 600 words for the 3rd year and from 800 to 1,000 words for the 4th year. The texts were saved in plain text format, so that they could be processed by the computing program *WordSmith Tools* (SCOTT, 2007). Later on, they were sent by e-mail to the researcher in order to organize them in folders. So far, the University Learners' Corpus has 102,608 words and it is aimed at reaching 250,000 words. The distribution of the corpus is in Table 2.

Table 2. University learners' corpus statistics.

Student university level	Sub-corpus size
1st year	2,638
2nd year	7,217
3rd year	22,004
4th year	70,749
Complete corpus	102,608

For this investigation, Mark Davies' *The Corpus of Contemporary American English* (COCA) (DAVIES, 2012) academic section was also used with the purpose of carrying out a comparative analysis with the collocational patterns extracted from the University Learners' Corpus. COCA is composed of 425 million words from more than 160,000 texts and its academic section consists of 76 million words[2].

Method

The methodological procedures were divided into two parts. The first one refers to the method used for finding out whether the teaching of collocations should be implicit or explicit to Brazilian university students. The second part will focus on the procedures for extracting collocaitons from the University Learners' Corpus and COCA academic section.

Methodological Issues for finding out whether the teaching of collocations should be implicit or explicit

First of all, students were given texts to be read and discussed, they did vocabulary exercises using the lexicon appeared in those texts, watched DVD series and movies with focus on the phraseological aspect and discussed the topics presented in these activities as well, without being drawn attention to the collocational aspects.

The themes were chosen according to their level of English. For example, one topic from the 2nd year of the B.A. in the English Language Teaching Course was *Smoking cigarettes* – text reading as homework assignment and a 2 hour-class on the

---

[2]COCA is freely available on the web at <http://corpus.byu.edu/coca/>, according to the references.

referred topic; another theme from the 3rd year of the Translation Course was *Stereotype* – 3 text readings as homework assignment and a 6 hour-class on the referred topic; and one topic from the 4th year of the Translation Course was *Capital Punishment* – 5 text readings as homework assignment and a 6 hour-class on the referred topic.

Later on, students were given a collocational test, focusing on some collocations which appeared in the texts, slides, videos, classroom discussions and vocabulary exercises. These tests (Appendix 1) were carefully prepared by the researcher in order to include only the collocations that were, in fact, studied in the classroom activities.

Collocation extraction from the university learners' corpus and COCA

The second part of the investigation presented in this paper focused on the extraction of the collocations from the University Learners' Corpus, the analysis of their lexicon-grammatical choices, and a comparative study of the collocational patterns found in the COCA.

To carry out this investigation, with the help of the computing program *WordSmith Tools* and its main tools *WordList*, *KeyWords* and *Concord*, it was firstly generated a wordlist for the texts from the University Learners' Corpus, sub-corpus of the 3rd and 4th year. Afterwards, using the wordlist of the reference corpus British National Corpus (BNC, 2012) and the wordlist of the Learner Corpus, a keyword list was generated. After analyzing this list, the keywords 'death penalty' and 'stereotype' were selected for this research.

The next step, now using the tool *Concord*, was to analyze the context in which these keywords occurred and extract some collocations. After doing that, the students were involved in a research using the online corpus COCA. They were supposed to look for collocations from the two bases investigated from their own essays ('death penalty' and 'stereotype'). As a result, they could discover a wider variety of possible collocates for the bases and the collocational patterns commonly or effectively employed by native speakers of English.

Results and discussion

Should the teaching of collocations be implicit or explicit?

As mentioned above, the learners were not drawn attention to the collocational aspects when they read the texts, did the exercises or any activities carefully prepared for the purpose of exposing them to this phraseological aspect of the language. After everything was carried out, discussed and done, they were given the collocational tests. They were not told to study for the test, but they were aware that they would be given a test to evaluate the vocabulary learnt in the last classes. It was also assured to them that this grade would not be included in the calculation of their grade average.

The result of the university learners' collocational tests demonstrated that most of them got very low marks, as presented in Table 3.

Based on Table 3, it can be seen that the students from the 2nd year of the B.A. in the English Language Teaching Course got the lowest marks, except for student 7, who got 9,0. This student has a proficient level of English and is more aware of the collocational aspect of the language. That also happened to student 5, from the 4th year of the Translation Course, who also has a proficient level. It may seem that as students become more proficient in the language (3rd and then 4th year), the collocational test result shows a rise in their grades.

Table 3. Learners' collocational test result.

English Language ii – Letters Course	Mark	English Language iii – Translation Course	Mark	English Language iv – Translation Course	Mark
Student 1	0.5	Student 1	4.5	Student 1	6.5
Student 2	1.0	Student 2	3.0	Student 2	5.0
Student 3	0.5	Student 3	7.5	Student 3	7.25
Student 4	0.5	Student 4	4.0	Student 4	3.75
Student 5	2.0	Student 5	4.0	Student 5	9.0
Student 6	1.5	Student 6	4.25	Student 6	4.5
Student 7	9.0	Student 7	5.0	Student 7	6.0
Student 8	1.0	Student 8	4.0	Student 8	4.0
Student 9	0.5	Student 9	6.0	Student 9	4.5
Student 10	1.0	Student 10	5.0	Student 10	4.5
Student 11	2.5	Student 11	6.0	Student 11	4.0
Student 12	0.5	Student 12	6.0	Student 12	7.0
		Student 13	6.5	Student 13	6.0
		Student 14	5.5		
		Student 15	4.5		
		Student 16	7.0		
		Student 17	3.0		
		Student 18	4.5		

Most students who got higher marks are more proficient than the others who did not so, except for student 16, from the 3rd year of the Translation Course who, in spite of being a more advanced student, got a low mark. That also occurred to the 4th year students, except from students 8 and 11.

It is worth mentioning this is the first phase of our experiment, as one future study aims at carrying out the same research with learners from next year's classes and expand it to other levels. It must be regarded that the number of subjects investigated has to be increased in order to have a more reliable result.

However, it is believed that the result of this partial work is increasingly meaningful as it may help prove that the teaching of collocations should not be implicit, but, on the contrary, that it should be intentional and explicit, so that students become more and more aware of the collocational and conventional aspect of the language. Based on our experience, after students were taught about the phraseological features of the English language, most of them reported to have never noticed or paid attention to such an aspect. Some learners even mentioned they thought they could use a particular collocate for any word they wanted to, as long as it made sense to them or it fit to that context, according to their view.

After they were shown their test results, some students have become more responsible for their learning and started reflecting on the possible collocates for the words studied in the classroom and become more careful about the combinability of words when producing their written or oral texts. Moreover, all of them recognized and stressed the importance of phraseology to the learning of a foreign language, as they realized that the research

proved they do not have problems regarding understanding the multi-word units, but they do have difficulties in producing them when speaking or writing. Hence, this investigation seems to have contributed to the development of such university students' collocational competence.

The analysis of collocations found in the university learners' corpus and COCA

For the extraction and analysis of the collocations and types of collocations employed by the Brazilian university learners, the keywords generated by the tool *Keywords* (SCOTT, 2007) were used. The first keyword raised from the Learners' Corpus investigated was 'death'. As it could be expected and verified in the corpus with the help of the program *Concord*, most examples had the word 'death' collocated with 'penalty', and hence forming the combination 'death penalty' (Figure 2).

Analyzing the concordance lines, the following collocations and types of collocations were extracted from the combination 'death penalty' (87 occurrences) (Table 4).

Figure 2. The first 30 concordance lines of 'death penalty'.

Table 4. Collocations and types of collocations extracted from the learner corpus.

Verbal collocations	Nominal collocations
Escape [the] death penalty	The practice of [the] death penalty
Deserve [the] death penalty	Alternatives to [the] death penalty
Get the death penalty	Result in the death penalty
Promote the death penalty	Methods of [the] death penalty
Use [the] death penalty	Legal administration of [the] death penalty

The symbol [ ] indicates that the learner omitted the definite article *the* when he should have used it.

When searched for the same combination in COCA, it was found 478 occurrences, out of which 87 concordance lines were analyzed and the following collocations extracted (Table 5).

According to Table 4, the university learners used only two types of collocations – 5 verbal collocations and 5 nominal collocations –, whereas in the first 87 concordance lines of COCA Academic (Table 5) it could also be found 6 adjectival collocations, besides 27 verbal collocations and 9 nominal collocations. On top of that, if we consider that the same number of concordance lines in the University Learners' Corpus and COCA (Academic Section) was analyzed, the number, frequency and variety of collocations extracted from the native speaker corpus is much higher than the number of collocations found in the University Learners' Corpus, even though the students had employed frequently used collocations. In addition, it is also possible to note that the collocations produced by the Brazilian learners involved in the research are very much the equivalent collocations to their mother tongue.

Another aspect that drew our attention to was the fact that the students did not use, as they should, the definite article 'the' in most collocations extracted with the basis 'death penalty', even though it had already been observed, in previous studies (ORENHA-OTTAIANO, 2010, 2011), an overuse of the definite article in Brazilian students' oral and written texts. According to data from the University Learners' Corpus, there were 87 occurrences of 'death penalty', out of which 10 collocations were extracted. However, out of these 10 collocations, the students correctly used the definite article in four of them. If we also observe the contexts with the combination 'death penalty', it is possible to notice that, out of the 87 occurrences, the learners used the definite article in only 6 occurrences (Figure 3).

Figure 3. Concordance line of 'the death penalty' from the learners' corpus.

On the contrary to what happened to the use of the definite article with the collocation death penalty', and asserting what it had previously been mentioned regarding Brazilian students' overuse of the definite article 'the' with the collocation 'death row', as shown below (Figure 4).

Table 5. Collocations and types of collocations extracted from COCA.

Verbal collocations	Nominal collocations	Adjectival collocations
Abolish the death penalty	Abolition of the death penalty	Applicable death penalty
Adopt the death penalty	Banning of the death penalty	First degree death penalty
Allow the death penalty	Condemnation of the death penalty	Juvenile death penalty
Apply the death penalty	Endorsement of the death penalty	The then-existing death penalty
Approve of the death penalty	Harshness of the death penalty	Unenforceable death penalty
Authorize the death penalty	Imposition of the death penalty	Worthy of the death penalty
Avoid the death penalty	Opposition to the death penalty	
Bring back the death penalty	Support of the death penalty	
Call for the death penalty	Unconstitutionality of the death penalty	
Defeat the death penalty		
End the death penalty		
Expand the death penalty		
Face the death penalty		
Impose the death penalty		
Invalidate the death penalty		
Oppose the death penalty		
Outlaw the death penalty		
Permit the death penalty		
Place a moratorium on the death penalty		
Prohibit the death penalty		
Reduce the death penalty		
Render the death penalty		
Restore the death penalty		
Retain the death penalty		
Seek the death penalty		
Spare the death penalty		
Turn away from the death penalty		

Figure 4. Concordance line of 'death row' from the learners' corpus.

The concordance lines show that, out of 8 occurrences of the collocation 'death row', only two were correctly used without the definite article 'the'. The same collocation with the definite article was searched in COCA (academic section). It was found 81 occurrences of 'death row' correctly employed without the definite article and no occurrences of this combination with the definite article. Searching it in Coca (all sections), 1876 examples could be found, out of which 36 had been used with the definite article 'the'. However, in all examples, 'death row' worked as an adjective as in 'death row inmates'. Hence, these findings help corroborate the overuse of the definite article 'the' in students' written texts. Moreover, it also shows the influence of the mother tongue in the use of the referred article.

The next keyword analyzed was 'stereotype'. Out of the 69 occurrences in the University Learners' Corpus, three types of collocations could be extracted: 4 verbal collocations, 3 nominal collocations and 3 adjectival collocations (Table 6).

Table 6. Collocations extracted from the node 'stereotype' (learners' corpus).

Verbal collocations	Nominal collocations	Adjectival collocations
Perpetuate a stereotype	The macho-man stereotype	Unfavorable stereotype
Develop a stereotype	The capitalism stereotype	Uncountable stereotype
Create a stereotype	The dumb blond stereotype	Historical stereotype
Change a stereotype		

Afterwards, the same basis was searched in COCA Academic, 762 occurrences were found and the following result of collocations was reached, grounded on the first 69 concordance lines analyzed (Table 7).

It can be noticed that 28 collocations were extracted from the 69 concordance lines analyzed and there were four types of collocations observed: 10 verbal collocations, 8 nominal collocations, 8 adjectival collocations and 2 adverbial collocations. Nevertheless, out of 69 occurrences with the node 'stereotype' found in the Learners' Corpus, the number, the frequency and the variety of collocations extracted were much lower: 4 verbal collocations, 3 nominal collocations, and 3 adjectival collocations. It was not found any example of adverbial collocations with the verb 'stereotype'.

Following the same methodology, other keywords ('deterrent', 'pain', 'crime' etc.) were analyzed and similar results were reached, which, however, they will not be discussed in the scope of this paper.

Table 7. Collocations extracted from the basis 'stereotype' (COCA).

Verbal collocations	Nominal collocations	Adjectival collocations	Adverbial collocations
Blow away a stereotype	Dope-smoking stereotype	Classic stereotype	Erroneously stereotype
Break a stereotype	Longhairs-whosmoke-dope stereotype	Cruel stereotype	Negatively stereotype
Challenge a stereotype	Male rock-star stereotype	Cultural stereotype	
Defy a stereotype	Gender stereotype	Deeply-rooted stereotype	
Disprove a stereotype	Sickly Negro stereotype	Inaccurate stereotype	
Fulfill a stereotype	Pop-culture stereotype	Massive stereotype	
Jump to a stereotype	The happy black sambo stereotype	Moronic stereotype	
Perpetuate a stereotype	The 'sugar-daddy' stereotype	Simpering stereotype	
Reinforce a stereotype			
Subside into a stereotype			

Final considerations

Even though conclusions should not be drawn hastily, taking into account that only the final results of the investigation may help us have a more well-grounded answer to the question posed in this article, the partial results of this research have shown to be particularly significant and may serve to stress that the teaching of collocations should be intentional and explicit. We strongly believe that when learners become more aware of the collocational and conventional aspect of the language, when they get to know the 'satellite words' that gravitate around the basic ones, and understand how a language works, they will be able to achieve

communicative mastery of English much faster and increase their academic achievement.

With respect to the extraction of collocations from the University Learners' Corpus and the comparative analysis of the referred collocations to the ones taken from the COCA Academic Section, the findings of this research suggested that, in this learning environment, learners have become more aware of collocational aspects, besides having become more responsible for their own learning as well. Although it could be noted that the number, the frequency, the type and the variety of collocations extracted from the University Learners' Corpus were lower than the ones found in COCA, a more careful analysis of the data should be carried out so that we could have a more accurate interpretation of the results.

The analyzed data led us to draw the tentative conclusion that Brazilian learners tend to use collocational patterns that are very similar to the equivalent patterns in their mother tongue. Furthermore, by taking part in the investigation and discussions, learners could notice that even though the texts from which the collocations were extracted are different (their essays and the ones from COCA Academic), there's been an underuse of collocations in their texts and that the analysis may indicate that the collocations produced by them are not so elaborate as expected for upper intermediate and advanced learners of English in the sense of trying to vary their writing and of offering alternative ways of saying things which may be more precise, according to the context. The research result was important to show that they can improve on their collocational production and, moreover, help them realize they have enough level of proficiency to use more sophisticated language – vocabulary and phraseological units.

To conclude, we hope to have shed some light on the development of students' collocational competence and shown the potential benefits of the collocation knowledge in English. However, the challenges for foreign language collocational patterns still remain: how can we minimize collocational errors, considering the difficulties learners have in mastering phraseological units and what type of teaching material should be created so that students achieve this goal? On the whole, we may notice that a lot has been done in terms of phraseological research; nevertheless, a lot still remains to be done if we really aim to achieve the goal of defining, extracting, describing, and contributing to the foreign language learning and the foreign language teaching pedagogy.

References

AIJMER, K. **Corpora and language teaching**. Amsterdam: John Benjamins, 2009.

ALTENBERG, B.; EEG-OLOFSSON, M. Phraseology in Spoken English: presentation of a Project. In: AARTS, J.; MEIJS, W. (Ed.). **Theory and practice in Corpus Linguistics**. Amsterdam: Randpi, 1990. p. 1-26.

BAHNS, J.; ELDAW, M. Should we teach EFL students collocations? **System**, v. 21, n. 1, p. 101-114, 1993.

BNC-British National Corpus. **Wordlist**. Available from: <http://www.lexically.net/downloads/version4/downloading%20BNC.htm>. Access on: Jan. 5, 2012.

CHANNELL, J. Applying semantic theory *to vocabulary teaching*. **ELT Journal**, v. 35, n. 2, p. 115-22, 1981.

CONZETT, J. Integrating collocation into a reading & writing course. In: LEWIS, M. (Ed.). **Teaching collocation**: further developments in the lexical approach. Croatia: Heinle, 2000. p. 70-87.

DAVIES, M. **The corpus of contemporary american english**: 425 million words, 1990-present. Available from: <http://corpus.byu.edu/coca/>. Access on: Fev. 20, 2012.

DE COCK, S.; GRANGER, S.; LEECH, G.; MCENERY, T. An automated approach to the phrasicon of EFL learners. In: GRANGER, S. (Ed.). **Learner english on computer**. London/New York: Longman, 1998. p. 67-79.

FONTENELLE, T. Towards the construction of a collocational database for translation students. **Meta XXXIX**, v. 39, n. 1, p. 47-56, 1994.

FRANKENBERG-GARCIA, A.; ASTON, G.; FLOWERDEW, L. **New trends in corpora and language learning (corpus and discourse)**. London: Continuum Publishing Corporation, 2011.

GRANGER, S. The International corpus of learner english. In: AARTS, J.; DE HAAN, P.; OOSTDIJK, N. (Ed.). **English language corpora**: design, analysis and exploitation. Amsterdam: Rodopi, 1993. p. 57-69.

GRANGER, S. **Learner english on computer**. London/New York: Longman, 1998.

GRANGER S.; HUNG, J.; PETCH-TYSON, S. **Computer learner corpora, second language acquisition and foreign language teaching**. Amsterdam/Philadelphia: John Benjamins, 2002.

HAUSMANN, F. J. Wortschatzlernen ist kollokationslernen. Zum lehren und lernen französischer Wortverbindungen. **Praxis des neusprachlichen Unterrichts**, v. 31, p. 395-406, 1984.

HAUSMANN, F. J. Kollokationen im deutschen wörterbuch. Ein beitrag zur theorie des lexikographischen Beispiels. In: BERGENHOLTZ, H.; MUGDAN, J. (Ed.). **Lexikographie und grammatik**. Tübingen: Niemeyer, 1985. p. 118-129.

HEID, U.; MARTIN, W.; POSCH, I. **Feasibility of standards for collocational description of lexical items**. Stuttgart: Eurotra - 7 Report, 1991.

HILL, J. Revising priorities: from grammatical failure to collocational success. In: LEWIS, M. (Ed.). **Teaching**

collocation: further developments in the lexical approach. Croatia: Heinle, 2000. p. 47-69.

MARTON, W. Foreign vocabulary learning as problem no. 1 of language teaching at the advanced level. **Interlanguage Studies Bulletin**, v. 2, n. 1, p. 33-57, 1977.

MEUNIER, F.; GRANGER, S. Phraseology in foreign language learning and teaching. Where to and from? In: MEUNIER, F.; GRANGER, S. (Ed.). **Phraseology in foreign language learning and teaching**. Amsterdam: John Benjamins, 2008. p. 247-252.

MEUNIER, F.; DE COCK, S.; GILQUIN, G.; PAQUOT, M. (Ed.). **A Taste for corpora**: In honour of Sylviane Granger. Amsterdam: John Benjamins, 2010.

NESSELHAUF, N. **Collocations in a learner corpus**. Amsterdam: John Benjamins, 2005.

ORENHA-OTTAIANO, A. **A compilação de um glossário bilíngüe de colocações, na área de jornalismo de negócios, baseado em corpus comparável**. 2004. 246f. Dissertação (Mestrado em Estudos Linguísticos e Literários)-Universidade de São Paulo, São Paulo, 2004.

ORENHA-OTTAIANO, A. **Unidades fraseológicas especializadas**: colocações e colocações estendidas em contratos sociais e estatutos sociais traduzidos no modo juramentado e não-juramentado. 2009. 282f. Tese (Doutorado em Estudos Linguísticos)– Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista, São José do Rio Preto, 2009.

ORENHA-OTTAIANO, A. **A tecnologia no ensino de línguas estrangeiras**: o uso do programa WordSmith Tools e de corpora on-line em inglês. Minicurso apresentado durante a XXII Semana de Letras do Instituto de Biociências, Letras e Ciências Exatas, da Universidade Estadual Paulista "Júlio de Mesquita Filho", set. 2010. São José do Rio Preto: Unesp, 2010.

ORENHA-OTTAIANO, A. **Corpora paralelos como ferramentas para a prática tradutória**. Minicurso apresentado durante a XXXI Semana do Tradutor do Instituto de Biociências, Letras e Ciências Exatas, da Universidade Estadual Paulista "Júlio de Mesquita Filho", set. 2011. São José do Rio Preto: Unesp, 2011.

PRAVEC, N. Survey of learner corpora. **ICAME Journal**, n. 26, p. 81-114, 2002.

SCOTT, M. **WordSmith tools**: version 5.0. Liverpool: Lexical Analysis Software Ltd., 2007.

SINCLAIR, J. **Corpus, concordance and collocation**. Oxford: Oxford University Press, 1991.

TONO, Y. Using learner corpora in ELT and SLA research. In: SYMPOSIUM ON THE ROLES OF CORPORA IN LANGUAGE TEACHING AND LANGUAGE ENGINEERING, 12th, 1999, Tokyo. **World Congress...** Tokyo: AILA, 1999.

Appendix 1

Sample of a collocational test

Part 1 – Fill in the blanks with a suitable word or phrase:

1. It's the job of the police to _____ the law.
2. There is to be a total _____ (= prohibition) on smoking in the office.
3. You'll have to _____ (= accept, obey) the rules of the club.
4. The first lethal injection in the country was _____ in that room in 1982.
5. Human trafficking is a _____ crime that is thriving in the United States.
6. He advised the organization on how to handle recruitment issues effectively _____ (=according to) the law and according to best practice.
7. Capital punishment in the United Kingdom was _____ in the 20th century (they don't have it anymore).
8. The manner in which the crime was _____ (= committed) indicates planning, sophistication, or professionalism.
9. The number of _____ (= big) crimes in the capital has gone down drastically as compared to last year thanks to effective policing. On the other hand, **10.** _____ (= small) crimes have risen alarmingly.