BELT+
Brazilian English Language Teaching Journal

ISSN 2178-3640

Porto Alegre, July-December 2018, v. 9, n. 2, p. 379-394

http://dx.doi.org/10.15448/2178-3640.2018.2.31799

ORIGINAL ARTICLE

Prepping a prep course: a corpus linguistics approach

Amanda Chiarelo Boldarine[1], Rodrigo Garcia Rosa[2]

[1] Faculdade Cultura Inglesa. São Paulo, SP, Brasil.
[2] Universidade de São Paulo (USP). São Paulo, SP, Brasil.

ABSTRACT

The purpose of this article is to explore and report some possible uses of corpus linguistics tools and techniques in a preparatory course for an international exam, focusing on helping students use corpora to find and analyze collocations and colligations when doing and creating multiple-choice cloze exercises. The participants were undergraduate students taking either a teacher training or a translation program. After discussing some research carried out on the pedagogical implications of using corpora, the article presents how a six-session preparatory course was designed and implemented, and the tools used to check participants' perception of learning. The compiled data analyzed how participants reacted towards using an online corpus while getting themselves ready for the Use of English component of exams. The results demonstrated that corpus techniques were felt to be useful tools as far as fostering students' autonomy was concerned. The study ends with a brief discussion on its limitations.

Keywords: corpus linguistics; teaching; multiple-choice cloze.

Criação de um curso preparatório: uma abordagem sob o viés da linguística de corpus

RESUMO

O objetivo deste artigo é explorar e reportar alguns possíveis usos de ferramentas de linguística de corpus em um curso preparatório para um exame internacional, focado em ajudar alunos universitários a usarem corpora para encontrarem e analisarem colocações e coligações, e a fazerem e criarem exercícios de multiple-choice cloze. Depois da introdução que discute algumas das pesquisas realizadas sobre as implicações pedagógicas do uso de corpora, apresentamos a estrutura do curso, que teve duração de seis sessões, como foi planejado e implementado, assim como as ferramentas utilizadas para verificar a percepção de aprendizagem dos alunos. O artigo conclui com os seguintes resultados: os alunos tiveram uma atitude positiva em relação ao uso de um corpus on-line, o que se evidenciou nas opiniões dos participantes ao salientarem a autonomia como um dos atributos trazidos pela utilização da ferramenta. Por fim, apresentamos uma breve discussão acerca de algumas limitações deste estudo.

Palavras-chave: linguística de corpus; ensino; multiple-choice cloze.

Corresponding Author:
Amanda Chiarelo Boldarine
<amanda_chiarelo@yahoo.com.br>

1. INTRODUCTION

The study of how Corpus Linguistics can contribute to language teaching dates back to the late 80's. Tim John's seminal work (John, 1986; 1991) focusing on the analyses of concordance lines done by learners, now known as *data-driven learning* (DDL), is commonly considered one of the first works to explore corpora for teaching purposes (Leech, 2013). Since then, however, a number of researchers have delved into the pedagogical implications of using corpora (Fligelstone, 1993; Gavioli & Aston, 2001; Braun, 2007; O'Keeffe, McCarthy, & Carter, 2007; Flowerdew, 2009; Römer, 2011; Berber Sardinha, 2011; Viana & Tagnin, 2011; Leech, 2013)[1].

Biber, Conrad and Reppen (1998) highlight that the importance of Corpus Linguistics lies in the fact that by using a corpus – "a large and principled collection of natural texts" (Biber, Conrad, & Reppen, 1998, p. 4) that is stored and processed by a software – one is able to analyze language empirically, since the analyst is able to study naturally occurring samples of written and spoken language. Therefore, when using corpora, students do not have to rely on their intuition or on a native speaker to explain and understand how language works, instead they can do this more autonomously (Berber Sardinha, 2011). Moreover, corpora might be used to provide learners with hands-on experience; this way, an inductive approach towards learning is adopted and learners have the chance to play the role of a researcher; as put forward by Tim Johns "every student is a Sherlock Holmes" (Johns, 2002 apud O'Keeffe, McCarthy, & Carter, 2007, p. 24).

In spite of the aforementioned advantages, the pedagogical implications of using corpora have not yet been fully explored in Brazil (Viana & Tagnin, 2011). In addition, Berber Sardinha (2011) claims that "it would be beneficial if we had more publications showing Brazilian teachers how to use corpora with Brazilian students in Brazilian schools/companies/lessons"[2] (p. 349). Hence, this paper aims at bridging this gap by providing a contribution to the field of Corpus Linguistics' application to teaching through the description and analysis of an experience in which corpus linguistics tools and techniques have been used in a preparatory course for an international exam (Certificate in Advanced English – CAE) with undergraduate students. Additionally, the experience herein described may hopefully serve as an empirical example of how Brazilian teachers can plan and deliver language courses with the aid of corpus linguistic tools and techniques.

This article is divided into five sections. The first section provides a brief rationale, context and the purpose of this study. The second section reviews the literature relevant to the topic and that guided our intervention. The third section describes the research methodology adopted by presenting the rationale for the course designed, the research participants and the tools. The fourth one presents the data gathered by the investigation. Finally, the fifth section evaluates and discusses the implications of the findings and indicates future research directions.

[1] See McEnery & Xiao (2010, p.364) for a thorough list of publications.
[2] Original:"Seria muito benéfico se tivéssemos mais publicações mostrando aos professores brasileiros como usar *corpora* com alunos brasileiros em escolas/empresas/aulas brasileiras".

2. CORPORA AND TEACHING

Sinclair (2013) advocates that teachers should, whenever possible, provide learners with real examples of language use and for that reason, corpus tools would be highly beneficial both for learners and teachers. Cook (1998), on the other hand, claims that corpora are tools that can certainly be used in teaching, but they should not be considered as the "only valid source of facts about language" (p. 58). The author shows concern about the negative impact that extreme views of Corpus Linguistics might have on teaching and questions whether a corpus could always be considered a reliable source of written and spoken language, as it does not account for the speaker/writer's communicative intentions. For him intuition and elicitation are also important tools when describing languages and corpora should be used to help describe languages and not prescribe what should be taught. On the use of intuition, Sinclair (2013) acknowledges its importance for teachers and learners, especially with regard to the analysis of words and sentences in isolation; however, the linguist stresses that multi-word units, and not only isolated words and sentences, abound in the structure of English and as such should be given top priority in teaching. The author states that access to these patterns is substantially facilitated with the use of corpus tools, since learners are given the opportunity to analyze real and more extended samples of language, facilitating their understanding of how words co-occur and how these language units convey meaning in context. Gavioli and Aston (2001) also point out that corpora could be used to test intuition and "help us make better-informed decisions" (ibid, p. 239). For the authors, the use of corpora makes it possible for "learners to problematize language, to explore texts, and to authenticate discourse independently and collectively" (ibid., p. 244).

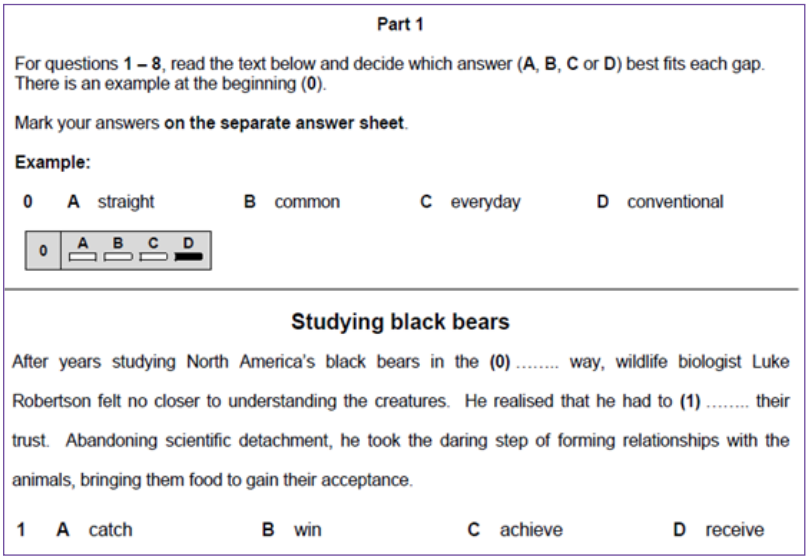
At a practical level, McEnery and Xiao (2010, p. 365) also mention a few constraints on the use of corpus tools such as the level and experience of learners, knowledge and skills required for corpus analysis and pedagogical mediation, and the access to resources. These are certainly challenging aspects for anyone who intends to introduce corpus tools into their teaching practice, but Corpus Linguistics enthusiasts and theorists have proposed different ways in which these challenges can be minimized. According to Leech (2013), corpora have been used in teaching contexts in, at least, two different ways: the direct and the indirect uses. For convenience of exposition we will start from the second one, which refers to an indirect use of corpus tools. This application, also mentioned by McEnery & Xiao (2010) as the most traditional use of corpus tools, refers to cases in which materials such as dictionaries, reference grammars, course books and even language tests make use of corpora as linguistic sources from which examples and descriptions of language patterns are drawn to be later tackled in the classroom. In other words, this indirect use is that in which the teaching materials are *corpus-informed* (Meunier & Reppen, 2015).

The first, and most frequently used way, refers to the direct use of corpora, for which Fligelstone (1993) sketches out some common steps on how it is usually implemented:

a) teaching about – which entails teaching about Corpus Linguistics as a field of study;

b) teaching to exploit – which is concerned with how one can use corpora; and
c) exploiting to teach – which means exploiting the resources offered by a corpus for teaching purposes.

Cobb & Boulton (2015) claim that the steps outlined above, especially teaching to exploit and exploiting to teach, can be truly beneficial for learners as they become more autonomous and independent language users. Learners can autonomously search for language patterns used in different genres and registers (Chong, 2013) as well as independently identify and select preferred and more conventional (and frequent) language patterns to be used in speaking and writing (Carter & McCarthy, 2006). As far as Leech's dichotomy is concerned, the perspective adopted in this study is to be categorized under the first group, that is, the one which uses corpora in a direct way. That said, as it will be clearer in section 3, the participants of this study were trained to use the corpus to observe *collocations*, the co-occurrence of words that have syntagmatic lexical relations (Crystal, 2008), and also *colligations*, the co-occurrence of lexical words with "grammatical markers or grammatical categories" (McEnery & Hardie, 2012, p. 130). The decision to target these two language structures had a general pedagogical reason as well as a practical one. The pedagogical reason had to do with the fact that being able to produce and process conventionalized units of language would be important for learners as "chunked expressions enable learners to reduce cognitive effort, to save processing time and to have language available for immediate use" (Shin & Nation, 2007, p. 340). The practical reason, on the other hand, is related to the kinds of language skills required from learners in international exams, like the Cambridge ESOL exams. In multiple-choice cloze exercises, for instance, candidates have to be able to understand the general idea of the text as well as identify collocations and colligations to choose the most appropriate option to complete the gaps, as can be seen in **Figure 1**:



Part 1

For questions **1 – 8**, read the text below and decide which answer (**A**, **B**, **C** or **D**) best fits each gap. There is an example at the beginning (**0**).

Mark your answers **on the separate answer sheet**.

Example:

0 **A** straight **B** common **C** everyday **D** conventional

0 | A | B | C | D

Studying black bears

After years studying North America's black bears in the **(0)** ..... way, wildlife biologist Luke Robertson felt no closer to understanding the creatures. He realised that he had to **(1)** ..... their trust. Abandoning scientific detachment, he took the daring step of forming relationships with the animals, bringing them food to gain their acceptance.

1 **A** catch **B** win **C** achieve **D** receive

Figure 1. Sample of a multiple-choice cloze exercise
Source: *Cambridge English Advanced– Handbook for teachers* (2016, p. 12).

As it is shown in **Figure 1**, learners should be able to identify that in (1), *'win'* collocates with *'trust'*, whereas the other options do not. In addition, collocations and colligations are pervasive in the exercise, not only in the gaps. For example, in *'he took the daring step of forming relationships with the animals'*, *'take the step'* or *'daring step'*, are possible collocations, as two lexical words co-occur in each expression, and *'relationships with'* can be an example of a colligation as a lexical word (*relationship*) co-occurs with a grammatical word (*with*)[3]. According to the *Handbook for Teachers* (2016), when doing the multiple-choice cloze exercises,

> Candidates are required to draw on their lexical knowledge and understanding of the text in order to fill the gaps. Some questions test at a phrasal level, such as collocations and set phrases. Other questions test meaning at sentence level or beyond, with more processing of the text required. A lexico-grammatical element may be involved, such as when candidates have to choose the option which fits correctly with a following preposition or verb form (p. 10, our highlights).

In light of the definition above, this type of exercise seemed to be the most conducive kind of language practice activity with which to use corpus tools since it tackles language aspects normally dealt with in Corpus Linguistics (collocations, fixed phrases, colligations, etc.). However, as it will be described in section 3.3, in this study we narrowed the application of corpus tools to collocations and colligations and left some other aspects out of the scope (set phases, for instance) due to time constraints. The course designed, then, aimed at equipping participants with the necessary tools to look for, identify and analyze collocations and colligations on an online corpus.

3. METHODOLOGICAL ISSUES: COURSE DESIGN AND IMPLEMENTATION

This section describes the methodology adopted in this study: how the course was designed and implemented, the participants, and the tools used.

3.1 Participants

The study was conducted at *Faculdade Cultura Inglesa* (São Paulo, Brazil) from March 6th 2018 to April 10th 2018, and students enrolled in two undergraduation courses, the Language Teaching Program (LTP) and the Translation Program (TP), were invited to participate. At *Faculdade Cultura Inglesa*, students are categorized according to their language proficiency for the English lessons; therefore, students taking the advanced course were invited and nine students volunteered[4]. Participants signed an informed consent as requested by *Faculdade Cultura Inglesa.*

[3] These collocations have been attested through a verification on COCA <https://corpus.byu.edu/coca/>. *'Take'* is the first verbal collocate to *'step'*, *'step'* is the twenty-sixth nominal collocate to *'daring'* and *'with'* is the second prepositional collocate to *'relationship'*.

[4] *Faculdade Cultura Inglesa* offers the subject 'Introduction to Corpus Linguistics' in its program, however, none of the participants in this study had taken it by the time this study was conducted.

Participants' first language is Brazilian Portuguese and a survey[5] was conducted in order to collect information regarding their educational background. **Table 1** summarizes the most relevant pieces of information.

Table 1. Summary of participants' profiles

	Gender	Years of English instruction	English Certificate	Program and year of study
Participant 1	Female	4 years	FCE	*LTP* – Second year
Participant 2	Female	3 years and a half	None	*LTP* – Fourth year
Participant 3	Female	10 years	None	*LTP* – Second year
Participant 4	Female	40 years	FCE	*LTP* – Third year
Participant 5	Male	Not mentioned	None	*LTP* – Third year
Participant 6	Female	7 years	FCE	*LTP* – Third year
Participant 7	Male	8 years and a half	None	*LTP* – Fourth year
Participant 8	Male	3 years	None	*TP* – First year
Participant 9	Female	Not mentioned	FCE	*LTP* – First year

3.2 The Corpus

For the purposes of this study, the *Corpus of Contemporary American English*[6] (COCA) was chosen due to its accessibility, ease of use, and because it is a morphosyntactically tagged corpus. COCA is one of the largest online corpora (containing 560 million words) available for free and it is divided into five genres: spoken, fiction, popular magazines, newspapers and academic journals; besides, 20 million words are added every year. It offers some search tools, such as, 'list', 'concordance', 'collocates' and KWIC (keyword in context). Participants were exposed to all the tools, but they were encouraged to use 'list' and 'collocates', as these tools would offer them more practical ways of comparing and contrasting the frequency of the collocates as well as checking the concordance lines (context). **Figure 2** below is an example of a PowerPoint slide used to show participants the importance of not only checking the frequency list, but also the context in order to verify collocates.

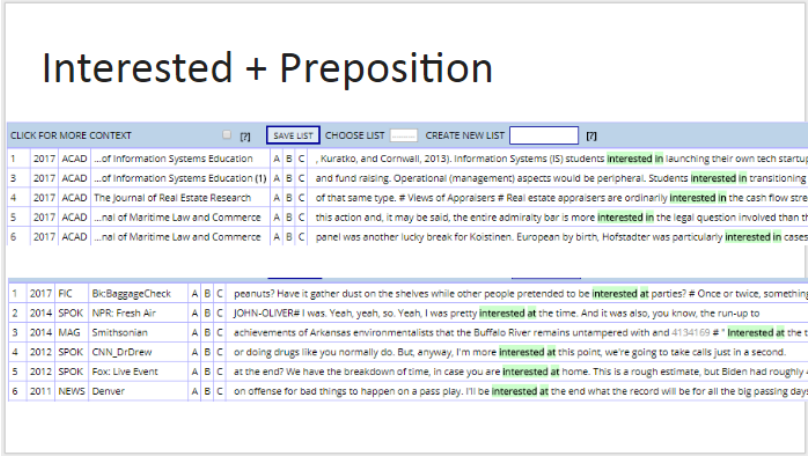


Figure 2. Slide used in the course (with screenshots of COCA for the search 'interested' + preposition)

[5] The survey can be found in **Appendix 1**.
[6] <https://corpus.byu.edu/coca/>.

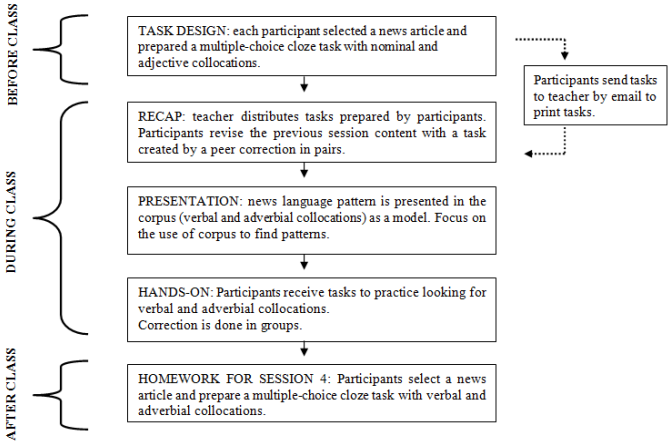
3.3 The Course

The course was divided into six 90-minute sessions offered as extra classes after participants' regular schedule. They were held at the computer lab, as COCA is only available online, participants needed to have access to computers and to the Internet; an interactive whiteboard was also available so the use of the corpus could be exemplified to the participants.

In the first session, participants did a needs analysis, which presented a multiple-choice cloze exercise taken from the CAE (Certificate in Advanced English) past paper. They were encouraged to do it without resorting to any kind of resources, such as the Internet or dictionaries. When correcting the exercise, participants were invited to justify their answers in order to check how linguistically aware they were with regard to their choices. Based on their justifications, the concept of collocations was presented and how a corpus could be useful to identify them. COCA was then introduced as well as how to create an account, login and use the 'list' tool.

First, a brief recap of the previous session was done; secondly, new language structures were presented (see **Table 2** below for the course syllabus) and how to search for some samples on COCA; thirdly, a hands-on moment was carried out to give participants the chance to search for the structures on COCA on their own. Finally, participants shared their hypotheses with peers and reported back their findings for a classroom discussion.

After the second session, as it was felt that participants were more familiarized with using COCA, they were assigned a piece of homework in which they had to select a text[7] and gap it (focusing on the same type of collocation/colligation discussed in the session) and check on COCA for possible answers and distractors, which should have been selected based on frequency and meaning. Participants would, then, do the multiple-choice cloze exercise created by their peers at the beginning of the subsequent session. Each session was meant to introduce, present and provide corpus practice on a different lexico-grammatical structure. Below is a schematized example of session 3 (**Scheme 1**) and following it **Table 2** brings the entire course syllabus.



Scheme 1. Example: Session 3.

[7] Participants were told to select texts from either <https://www.theguardian.com/uk or from https://www.nytimes.com/>.

Table 2. Course syllabus

Session	Content
First	Needs analysis + feedback Introduction to corpus linguistics tools
Second	Nominal collocation: Noun + Noun; Noun + Prep + Noun – e.g.: *hate crime* Adjective collocation: Adj + Noun – e.g.: *fat chance*
Third	Verbal collocation: Verb + Noun – e.g.: *make a mistake* Adverbial collocation: Adv + Verb; Adv + Adj – e.g.: *wholeheartedly agree; deeply offended*
Fourth	Colligation: Noun + Prep; Adj + Prep – e.g.: *insight into; interested in*
Fifth	Colligation: Verb + Prep – e.g.: *hope for*
Sixth	Wrap-up + questionnaire

Prior to the sixth session, participants sent their homework assignment[8] by email, so that their exercise could be checked and printed by the teacher; however, participants were told the language aspect that they had to focus on (two adverbial collocations; two nominal collocations; two verbal collocations; two colligations). In the session, participants did two exercises from the ones they had produced, and they helped each other to correct them. There was a wrapping-up discussion about creating and doing the multiple-choice cloze exercise, and they answered a questionnaire, which is discussed in the following section.

3.4 Questionnaire

The aim of using a questionnaire was to access participants' attitude towards the use of COCA and their perception of learning. The questionnaire[9] was designed by following the steps outlined by Dörnyei & Csizér (2012). A *Likert scale* was used, with close-ended items which participants had to rate from 'strongly disagree' to 'strongly agree'. First, three focal areas were selected to be investigated in the survey:

- how using COCA would raise their perception of language awareness;
- how using COCA would equip them with tools to become more autonomous language learners;
- how using COCA would better prepare them to sit for the Cambridge English: Advanced.

Second, the items covering those areas were written, following the guidelines proposed (ibid, p. 76), that is, "aim at short and simple items; use simple and natural language; avoid ambiguous or loaded words or sentences; avoid negative constructions; avoid double-barreled questions". Third, some personal questions were included at the end of the questionnaire. Dörney & Csizér (2012) claim that questionnaires tend to be more accurately answered when done in the participants' first language; however, due to the level of proficiency of the participants of this study and the fact that all instruction was carried out in English, we decided not to write the questionnaire in Portuguese.

[8] Assignment question: Choose a text and select 8 gaps, two of each have to be nominal collocations, adverbial collocations, verbal collocations and colligations. Remember to use COCA to check the alternatives.
[9] See **Appendix 1**.

4. DATA ANALYSIS: QUESTIONNAIRE'S ANSWERS AND PARTICIPANTS' TASKS

This section focuses on the data gathered in the study. Five of the exercises created by the participants for the last session were selected. In addition to the questionnaire, which aimed at verifying whether or not their perception of their use of COCA was positive, participants' production was also analyzed in an attempt to check the quality of their production and to what extent their perception of the use of the tool and their production matched.

4.1 *Likert* scale results

Table 3 shows the items of the questionnaire participants answered and also the focal area of each item. **Table 4** below depicts participants' answers in percentages. Overall, participants seemed to have a positive attitude towards using COCA, as most of the answers are on the 'agree' edge of the scale. A graph with the percentage of answers on the *Likert* scale (**Figure 3**) can also be found below.

Table 3. Questionnaire items and focal areas

Item	Focal Areas
1. Using COCA has helped me be better able to do multiple-choice cloze exercises.	Exam preparation
2. Using COCA has helped me create multiple-choice cloze exercises.	Exam preparation
3. Using COCA has helped me identify collocations and colligations.	Language awareness
4. Using COCA has helped me better understand how English is used.	Language awareness
5. I have started to use COCA to plan lessons.	Autonomy
6. I have started to use COCA to write texts.	Autonomy
7. I have started to use COCA to check my hypotheses of collocations and colligations.	Autonomy
8. I have started to spot more collocations and colligations when reading texts.	Language awareness
9. I have started to spot more collocations and colligations when watching films/videos.	Language awareness
10. I have started to pay more attention to my use of collocations and colligations when writing.	Language awareness
11. I have started to pay more attention to my use of collocations and colligations when speaking.	Language awareness
12. I feel I am better prepared to sit for CAE.	Exam preparation
13. I feel I am better prepared to do language analyses.	Autonomy
14. I feel I am better prepared to use COCA.	Autonomy

Table 4. Questionnaire result. Participants' answers to the questionnaire in percentages

Item	Total	Strongly agree	Agree	Partly agree	Slightly disagree	Disagree	Strongly disagree	Total
Item 1	**9**	33%	44%	22%	0%	0%	0%	**100%**
Item 2	**9**	56%	33%	11%	0%	0%	0%	**100%**
Item 3	**9**	22%	44%	33%	0%	0%	0%	**100%**
Item 4	**8**	22%	56%	11%	0%	0%	0%	**89%**
Item 5	**9**	0%	33%	56%	0%	11%	0%	**100%**
Item 6	**9**	33%	44%	0%	22%	0%	0%	**100%**
Item 7	**9**	67%	11%	22%	0%	0%	0%	**100%**
Item 8	**8**	22%	33%	22%	0%	11%	0%	**89%**
Item 9	**9**	11%	33%	33%	22%	0%	0%	**100%**
Item 10	**9**	56%	22%	22%	0%	0%	0%	**100%**
Item 11	**9**	11%	22%	33%	33%	0%	0%	**100%**
Item 12	**9**	11%	44%	44%	0%	0%	0%	**100%**
Item 13	**9**	22%	33%	22%	22%	0%	0%	**100%**
Item 14	**9**	89%	11%	0%	0%	0%	0%	**100%**

Considering participants' perception of language awareness, it can be concluded that using COCA might have been beneficial, which can be seen in items 4, 8 and 10. As for autonomy, in item 7, 67% of the answers were 'strongly agree'; in item 6, 33% were 'strongly agree' and 44 % 'agree'; and in item 13 most of the answers are on the 'agree' edge of the scale. Based on these results, it can be assumed that participants started using COCA to check their hypotheses of collocations and colligations. However, as seen in items 5, participants may not have started using COCA as a tool to help them plan lesson, as 56% of the answers were 'partly agree'.

Participants' perception of using COCA to get prepared for the CAE exam was also positive; for example, in item 12, 11% of the answers were 'strongly agree' and 44% were 'agree'. In addition, in items 1 and 2 participants' answers seem to show that using COCA has helped them do and create multiple-choice cloze exercises.

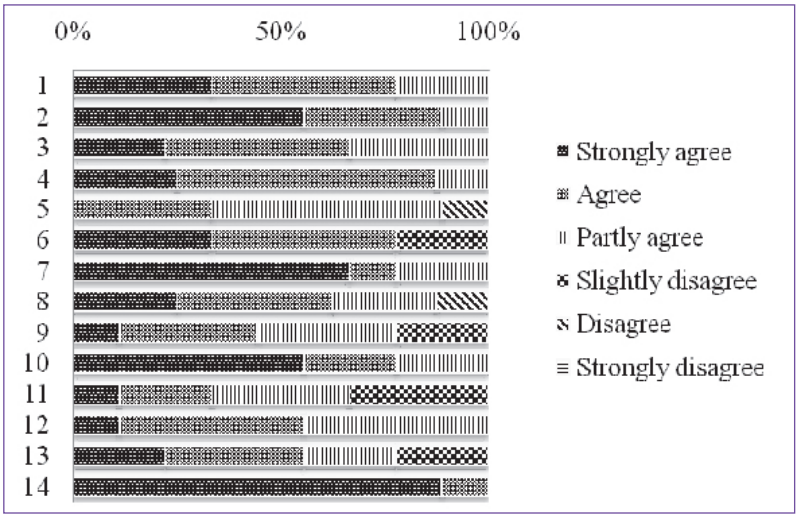


Figure 3. *Likert* scale results. Participants' responses are represented by different patterns and expressed in percentages and the numbers on the left side of the bars represent the items.

4.2 Participants' multiple-choice cloze tasks

This part is dedicated to the analyses of participants' production for the sixth session. Participants were asked to choose an article and gap 8 sentences. They had to come up with three alternatives that would be distractors by checking on COCA, based on frequency and meaning. Here, due to space constraints, we will show only a few selected sentences (not the entire texts) to be exemplified and analyzed.

Production of participant 1[10]

(1) This week, that democracy may be_____ eroded as a three-judge appellate court decides whether the most popular political figure in the country (…)
a. severely b. significantly c. so d. further

[10] The sample sentences were taken from an article published by *The New York Times* available on <https://www.nytimes.com/2018/01/23/opinion/brazil-lula-democracy-corruption.html>.

(2) He is accused of having accepted a bribe from a big construction company, called OAS, which was prosecuted in Brazil's "Carwash" corruption _____.
a. system b. scheme c. arrangement d. investigation

(3) There is not much pretense that the court will be impartial. The presiding judge of the appellate panel has already _____ the trial judge's decision to convict Mr. da Silva for corruption (…)
a. glorified b. praised c. celebrated d. exalted

(4) (…) the Workers' Party government _____ autonomy to the judiciary to investigate and prosecute official corruption (…)
a. granted b. gained c. acquired d. declared

Analyzing the multiple-choice cloze exercise created by this participant, it can be concluded that she relied on COCA to check the collocations and the colligations gapped. For example, in (1) '***further*** *erode'*, *'severely'* and *'significantly'* were selected as options and these two were found as collocates of *'erode'* as well, although not as frequently as *'further'*.

It was also observed that this participant was careful when selecting the options, as not only was frequency a used criterion, but also meaning. Some of the options selected were not found as collocates on COCA, but it could be for not having other options that could keep the same meaning as the collocation chosen in the text – it seems that the participant decided to opt for near synonyms. For instance, in (2)*'**corruption** scheme'* two of the options appear on COCA *'investigation'* and *'scheme'*, but the other two options *'system'* and *'arrangements'* do not. Similarly, in (3)*'**praised** the decision'*, she offered *'glorified'* and *'exalted'*. On the other hand, at times the participant may not have identified a whole chunk as in (4)*"**granted** autonomy to"* the participant selected as options: *'gain'*, *'acquire'*, and '*declare'* which would not be appropriate (one does not *gain/acquire/declare* autonomy *'to'* someone). It can be inferred that the preposition *'to'* might have been overlooked.

Production of participant 2[11]

(5) (…) adding that the impact of random chance on language had not been _____ appreciated before.
a. fully b. entirely c. minimally d. barely

(6) But a new study shows that another evolutionary mechanism might play a key _____: random chance.
a. part b. role c. function d. game

Once again, it seems that the participant intended to keep the meaning of the collocates, which can be seen in (5)*'**fully** appreciate'*; here the options given were '*barely*', '*entirely*' and *'minimally'*. Even though, '*entirely'* and '*minimally'* do not collocate with *'appreciate'*, it might be likely that the participant attempted to provide options that would virtually carry the same meaning as the other two possible collocates '*fully*' and '*barely*'. However, sometimes

[11] The sample sentences were taken from an article published by *The Guardian* available on <https://www.theguardian.com/science/2017/nov/01/resistance-to-changes-in-grammar-is-futile-say-researchers>.

resorting to synonyms was not done appropriately. In (6) *'play a key **role**'*, the options given were *'part'*, *'function'* and *'game'* and except for *'part'*, the other options are not possible collocates. It can be hypothesized that *'function'* and *'game'* were offered, because they can be found as synonyms for *'role'*[12], at the expense of the meaning in context and/or the collocation.

Production of participant 3[13]

(7) They have highlighted how decades of ____ cuts to public education have ____ these flames.
a. funding b. wage c. government d. neoliberal
a. spewed b. extinguished c. kindled d. deflected

(8) They have talked about how stagnant salaries mean teachers are unable to keep____ with the rising cost of healthcare.
a. tabs b. pace c. faith d. score

This participant seems to have relied on COCA to create the multiple-choice cloze exercise; the chosen alternatives for the gaps were not only based on frequency, but also based on meaning. For instance, in (7)*'**kindle** these flames'* the options offered were *'spew'*, *'extinguish'* and *'deflect'* and the four options are not frequent. It can be concluded that the participant was aiming at checking meaning at sentence and text level, as it was required to understand the text in order to choose the most appropriate answer. Another example would be in (8)*'keep **pace** with'* and the options given were *'tabs'*, *'faith'* and *'score'*, although *'score'* and *'tabs'* are very infrequent.

At times, the collocates chosen by the participant were not found on COCA, for example in (7), *'**neoliberal** cuts'*, but the options were, in this case *'funding'*, *'wage'* and *'government'*. Taking into consideration the fact that the exercises were created to resemble the multiple-choice cloze from CAE, it is likely that the collocation chosen would not be appropriate.

Production of participant 4[14]

(9) They are trying to figure out how to strike a ____ of engaging in Dr. King's (…)
a. match b. blow c. deal d. balance

(10) But presiding over a Pentecostal denomination of roughly six million members worldwide, he is well ____of his power.
a. off b. aware c. past d. intentioned

When verifying the collocations and colligations chosen by this participant, it was noticed that COCA was used and that the participant favored frequency over meaning. In six out of eight collocations and/or colligations selected, the options provided were at the top of the frequency list, regardless of the meaning conveyed by the collocation or the main idea of the text.

[12] It was found on <https://www.thesaurus.com/browse/role?s=t>.
[13] The sample sentences were taken from an article published by *The Guardian* available on <https://www.theguardian.com/commentisfree/2018/apr/10/women-teachers-strikes-america>.
[14] The sample sentences were taken from an article published by *The New York Times* available on <https://www.nytimes.com/2018/04/03/us/mlk-church-civil-rights.html>.

For instance, in (9)*'strike a **balance**'* the options selected were *'match'*, *'blow'* and *'deal'*, which are frequently preceded by 'strike', however, they would not be competing collocates when taking into account the text that was about Martin Luther King.

Although this participant seems to have become familiarized with using COCA and its tools, what might have prevented her from creating a more appropriate exercise could be a possible lack of awareness of collocations and colligations. In (10)*'well **aware** of'*, the options given were *'off'*, *'past'* and *'intentioned'*, which collocate with *'well'* but are not followed by *'of'* and the meaning conveyed by them would not fit the context.

Production of participant 5[15]

(11) Chilled, however, it tasted nicer than ___ nut milk I've bought: a thick water with a fleshed-out peanut flavor.
a. any b. plenty c. many d. lots

(12) Peanut milk is the ___ addition to the canon of nut milks that includes coconut, cashew, hazelnut and almond.
a. plenty b. healthy c. famous d. latest

One of the items that was marked as 'slightly disagree' was the one asking if the participant felt better prepared to do language analyses and that is aligned with what can be observed from the multiple-choice cloze created. Moreover, it is unlikely that COCA was used for creating the multiple-choice cloze exercise – most of the collocations and the colligations chosen were not found on the corpus.

Analyzing the options and the chunks selected, it may be inferred that the participant may have not understood how to identify collocations and colligations; for example (11), *'**any** nut milk'* was chosen as a collocation and the alternatives for *'any'* were *'plenty'*, *'many'* and *'lots'*. Another fact that might have prevented this participant from identifying collocations and using COCA may be a relative lack of language awareness; in (12) *'the **latest** addition'*, the alternatives chosen were *'powerful'*, *'healthy'* and *'famous'*, thus it can be concluded that the participant has failed to identify the superlative structure.

4.3 Limitations of the study

Most participants chose the collocates based not only on frequency, but also on meaning (as seen in the exercise created by participants 1 and 3), whereas some participants did not, for example, participants 4 and 5. It could be inferred that these participants might have failed to grasp the context in which the collocations and colligations occurred, which might corroborate the claims made by Braun (2007) that as most corpora are not built for teaching purposes, some students might find it difficult to retrieve context, or they might not have fully understood how to identify collocations

[15] The sample sentences were taken from an article published by *The Guardian* available on <https://www.theguardian.com/lifeandstyle/2018/apr/05/could-peanut-milk-be-new-star-of-nut-milk-scene>.

and colligations. However, the study serves the purpose of delimiting the challenges to be tackled in future corpus interventions.

It can also be envisaged that these participants would have benefitted from more consistent feedback on their performance and, as suggested by Gavioli & Aston (2001), from more moments in which they would have shared and compared their hypotheses and analyses, working in mixed-ability and mixed-level groups. Working more in mixed groups would have been beneficial for students, once "in this scaffolding-type of activity, more proficient students were able to offer their insights and interpretations on the corpus data, thus assisting the weaker students to gradually develop more independence" (Flowerdew, 2009, p. 404).

5. FINAL CONSIDERATIONS

The findings in this study suggest that participants are likely to have a positive attitude towards using corpora. Also, they have started using COCA to check their hypotheses of collocations and colligations, thus it can be profitable to exploit corpora to help participants become more autonomous and provide them with tools to test their intuition, as suggested by Berber Sardinha (2011) and O'Keeffe, McCarthy and Carter (2007). However, as pointed out by McEnery and Xiao (2010), students' level of proficiency may prevent them from making the most of corpora, which can be seen in the production of participant 5.

In conclusion, teaching to exploit corpora and exploiting them to teach may make students aware of collocations and colligations as well as make them more autonomous. However, when planning the lessons, the type of interaction among students should be carefully thought through and feedback moments should be more consistent, mainly when creating multiple-choice cloze exercises, as it requires different skills related to assessment and evaluation. It is also believed that in future studies other aspects of the language could also be dealt with in the sessions, for example, register, phraseologies, semantic prosody and idioms, which are also tested in C1 level exams.

REFERENCES

Berber Sardinha,T. 2011. Como usar a Linguística de *Corpus* no ensino de língua estrangeira – por uma política de Corpus Nacional Brasileira. In V. Viana & S. Tagnin (Orgs.). *Corpora no ensino de línguas estrangeiras* (p. 301-356). São Paulo: Hub Editorial.

Biber, D., Conrad, S., & Reppen, R. 1998. *Corpus Linguistics investigating language structure and use*. Cambridge: Cambridge University Press.

Braun, S. 2007. Integrating corpus work into secondary education: from data-driven learning to needs-driven corpora. *ReCALL*, 19(3), p. 307-328.

Carter, R. & McCarthry, M. 2006. *Cambridge grammar of English – a comprehensive guide*. Cambridge: Cambridge University Press.

Chong, C. 2013. 5 ways of using corpora to develop learner autonomy. *English Teaching Professional*, [Blog Post] August. Available on: <https://www.etprofessional.com/5-ways-of-using-corpora-to-develop-learner-autonomy>. Accessed on: 05 Nov. 2018.

Cobb, T. & Boulton, A. 2015. Classroom applications of corpus analysis. In D. Biber & R. Reppen (Eds.). *The Cambridge handbook of English corpus linguistics* (p. 478-496). Cambridge: Cambridge University Press.

Cook, J. 1998. The uses of reality: a reply to Ronald Carter. *ELT Journal,* 52(1), p. 57-63.

Crystal, D. 2008. *A dictionary of Linguistics and Phonetics.* Blackwell Publishing.

Dörnyei, Z. & Csizér, K. 2012. How to design and analyze surveys in second language acquisition research. In Alison Mackey & Susan M. Gass (Orgs.). *Research methods in second language acquisition: a practical guide* (p. 74-94). Wiley-Blackwell.

Fligelstone, S .1993. Some reflections on the question of teaching, from a corpus linguistics perspective. *ICAME Journal* 17, p. 97-109.

Flowerdew, L. 2009. Applying corpus linguistics to pedagogy – a critical evaluation. *International Journal of Corpus Linguistics,* 14(3), p. 393-417.

Gavioli, L. & Aston, G. 2001. Enriching reality: language corpora in language pedagogy. *ELT Journal*, 55(3), p. 238-246.

Handbook for teachers for exams from 2016. Available on: <http://www.cambridgeenglish.org/images/167804-cambridge-english-advanced-handbook.pdf>. Accessed on: 03 Feb. 2018.

Johns, T. 1986. Microconcord: a language-learner's research tool. *System*, 14(2), p. 151-162.

Johns, T. 1991. Should you be persuaded – two samples of data-driven learning materials. In T. Johns & P. King (Orgs.). *ELR Journal*, 4, p. 1-16.

Leech, G. 2013. Teaching and language corpora: a convergence. In. A. Wichmann, S. Fligelstone, T. McEnery & G. Knowles (Orgs.), *Teaching and language corpora.* London & New York: Routledge. E-Book. ISBN 13: 978-0-582-27609-3.

McEnery, T. & Hardie, A. 2012. *Corpus Linguistics*. Cambridge: Cambridge University Press.

McEnery, T. & Xiao, R. 2010. What corpora can offer in language teaching and learning. In E. Hinkel (Ed.). *Handbook of research in second language teaching and learning* (Vol. 2, p. 364-380). London & New York: Routledge.

Meunier, F. & Repper, R. 2015. Corpus versus non-corpus-informed pedagogical materials: Grammar as focus. In D. Biber & R. Reppen (Eds.). *The Cambridge handbook of English Corpus Linguistics* (p. 498-514). Cambridge: Cambridge University Press.

O'Keeffe, A., McCarthy, M., & Carter, R. 2007. *From corpus to classroom: language use and language teaching*. Cambridge: Cambridge University Press.

Römer, U. 2011. Corpus research applications in second language acquisition. *Annual Review of Applied Linguistics*, 31, p. 205-225.

Shin, D. & Nation, P. 2007**.** Beyond single words: the most frequent collocations in spoken English. *ELT Journal*, 62(4), p. 339-348.

Sinclair, J. M 2013. Corpus evidence in language description. In. A. Wichmann, S. Fligelstone, T. McEnery & G. Knowles (Orgs.). *Teaching and language corpora.* London & New York: Routledge. E-Book. ISBN 13: 978-0-582-27609-3.

Viana, V. & Tagnin, S. 2011. Introdução. In V. Viana & S. Tagnin. *Corpora no ensino de línguas estrangeiras* (p. 19-23). São Paulo: Hub Editorial.

APPENDIX 1 – SURVEY

We would like to ask you to help us by answering the following questions concerning the sessions. As this is not a test, there are no "right" or "wrong" answers and it is not necessary to write your name on it. Please give your answers sincerely in order to guarantee the success of the investigation. Thank you very much for your help.

I. Read the statements below carefully. After each statement you will find six boxes; please, put an 'X' in the box which best expresses the extent to which you agree with the statement.

	Strongly disagree	Disagree	Slightly disagree	Partly agree	Agree	Strongly agree
1. Using COCA has helped me be better able to do multiple-choice cloze exercises.						
2. Using COCA has helped me create multiple-choice cloze exercises.						
3. Using COCA has helped me identify collocations and colligations.						
4. Using COCA has helped me better understand how English is used.						
5. I have started to use COCA to plan lessons.						
6. I have started to use COCA to write texts.						
7. I have started to use COCA to check my hypotheses of collocations and colligations.						
8. I have started to spot more collocations and colligations when reading texts.						
9. I have started to spot more collocations and colligations when watching films/videos.						
10. I have started to pay more attention to my use of collocations and colligations when writing.						
11. I have started to pay more attention to my use of collocations and colligations when speaking.						
12. I feel I am better prepared to sit for CAE.						
13. I feel I am better prepared to do language analyses.						
14. I feel I am better prepared to use COCA.						

II. Finally, please answer a few personal questions.

15. How long have you been studying English? _____
16. Have you got an English certificate? Please underline: **YES** **NO**
17. If so, please write which one and when you sat for it: _____
18. How did you get prepared for it? _____
19. What do you study? _____
20. How long have you been studying this subject? _____

Thank you very much – we really appreciate your help!

Submetido: 30/08/2018
Aceito: 08/11/2018