

UNIVERSIDADE FEDERAL DE MINAS GERAIS

FACULDADE DE LETRAS

PROGRAMA DE PÓS-GRADUAÇÃO EM ESTUDOS LINGÜÍSTICOS

KATHERINE NUNES PEREIRA OLIVA

INTEGRATING CORPORA WITH LANGUAGE CLASSROOM:

ONLINE CORPUS AS AN EDITING TOOL

BELO HORIZONTE

2018

**FACULDADE DE LETRAS DA UFMG
KATHERINE NUNES PEREIRA OLIVA**

INTEGRATING CORPORA WITH LANGUAGE CLASSROOM:

ONLINE CORPUS AS AN EDITING TOOL

Dissertação apresentada ao Programa de Pós-Graduação em Estudos Linguísticos da Faculdade de Letras da Universidade Federal de Minas Gerais, como requisito parcial para obtenção do título de Mestre em Linguística Aplicada.

Área de Concentração: Linguística Aplicada

Linha de Pesquisa: Ensino/Aprendizagem de Línguas Estrangeiras – 3A

Orientadora: Profa. Dra. Deise Prina Dutra

**Belo Horizonte
Faculdade de Letras da UFMG
2018**

Ficha catalográfica elaborada pelos Bibliotecários da Faculdade de Letras/UFMG

O48i

Oliva, Katherine Nunes Pereira.

Integrating corpora with language classroom [manuscrito] :
online corpus as na editing tool / Katherine Nunes Pereira Oliva.
– 2018.

123 f., enc. : il., tabs., grafs., color.

Orientadora: Deise Prina Dutra.

Área de concentração: Linguística Aplicada.

Linha de pesquisa: Ensino/Aprendizagem de Línguas
Estrangeiras.

Dissertação (mestrado) – Universidade Federal de
Minas Gerais, Faculdade de Letras.

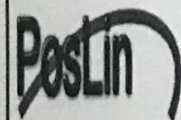
Bibliografia: f. 60-66.

1. Linguística de corpus – Teses. 2. Língua inglesa –
Métodos de ensino – Teses. 3. Língua inglesa (Ensino superior)
– Estudo e ensino – Teses. 4. Língua inglesa – Gramática –
Estudo e ensino – Teses. 5. Aquisição da segunda linguagem –
Teses. 6. Redação acadêmica – Teses. I. Dutra, Deise Prina. II.
Universidade Federal de Minas Gerais. Faculdade de Letras. III.
Título.

CDD : 420.7



UNIVERSIDADE FEDERAL DE MINAS GERAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTUDOS LINGÜÍSTICOS



FOLHA DE APROVAÇÃO

**INTEGRATING CORPORA WITH LANGUAGE CLASSROOM:
 ONLINE CORPUS AS AN EDITING TOOL**

KATHERINE NUNES PEREIRA OLIVA

Dissertação submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em ESTUDOS LINGÜÍSTICOS, como requisito para obtenção do grau de Mestre em ESTUDOS LINGÜÍSTICOS, área de concentração LINGÜÍSTICA APLICADA, linha de pesquisa Ensino/Aprendizagem de Línguas Estrangeiras.

Aprovada em 02 de julho de 2018, pela banca constituída pelos membros:

Prof(a). Bárbara Malveira Orfanó – Presidente da banca/Representante da orientadora
 UFMG

Prof(a). Valdênia de Carvalho e Almeida
 UFV

Prof(a). Climene Fernandes Brito Arruda
 UFMG

Belo Horizonte, 2 de julho de 2018.

À minha família

Divalda, minha mãe querida, exemplo de ser humano, conselheira de todas as horas, amor e dedicação incondicional; meu pai José Henrique, similar de dignidade e perseverança; meu irmão Henrique Oliva, pelo carinho e confiança depositada. Muitos contribuíram para a continuidade e dedicação aos estudos até a conclusão do Mestrado.

ACKNOWLEDGEMENTS

“Aqueles que passam por nós, não vão sós, não nos deixam sós. Deixam um pouco de si, levam um pouco de nós.”

Antoine de Saint-Exupéry

À minha orientadora, professora Deise Prina Dutra, que durante meus anos na graduação me acompanhou como professora, coordenadora e agora orientadora. Ao ser coordenada por ela no programa Idiomas Sem Fronteiras, despertei para o mundo da pesquisa acadêmica, o prazer de estar envolvida em projetos na universidade, a curiosidade pela Linguística de Corpus. Por indicação dela, hoje tenho meu certificado CELTA, um grande motivador para essa pesquisa na busca de melhorar as formas de ensino e aprendizado. Admiro imensamente a professora Deise como profissional e como pessoa. Agradeço pela paciência no decorrer deste trabalho, não foi fácil, muitos obstáculos apareceram no caminho e tive apoio e colaboração em todos os momentos de dificuldade e insegurança. Deise, em você me inspiro, admiro e reconheço o grande valor. Muito obrigada, não há agradecimentos que cheguem; sou eternamente grata.

Aos integrantes do Grupo de Pesquisa *Learner Corpus Research Group*, por estarem sempre presentes dando dicas, ouvindo, participando e me ajudando a crescer.

À Ilustre professora Bárbara Orfano pela assistência na pesquisa e aplicação em sala de aula.

Ao amigo Rodrigo, pela amizade, disposição em ajudar. Sempre pronto para me socorrer. Agradeço imensamente por todos os momentos se mostrar presente e compartilhar comigo seus conhecimentos sem hesitar. Seu apoio e parceria foram fundamentais nessa caminhada.

À amiga Raquel Rossini, pelo apoio, amizade e contribuição na evolução dessa pesquisa.

À amiga Andressa, pelo auxílio e despertar em mim o interesse por estudos com corpus e incentivo para o ingresso no mestrado, também a quem inspiro e tenho infinita gratidão.

À Ana Lúcia e Albina, por terem ensinado que existem mil e um meios de aplicar os conhecimentos durante o CELTA, peças-chaves para o tema desta pesquisa.

Ao Mateus, pelo apoio e companheirismo durante a trajetória da pesquisa e estímulo a superar esta jornada.

À Raissa, pelos debates e motivações em continuar sempre estudando e batalhando.

Ao Bernardo, pelo carinho, amizade e atenção durante os anos de faculdade até hoje.

À Bela, por encher de alegria os dias de insegurança.

Ao Bruno, pelo apoio e zelo em momentos de fraqueza.

Ao Ismael, pelo cuidado, amizade e doação, contribuindo para meu crescimento e fortalecimento nessa trajetória.

Aos colegas de trabalho do Colégio Santo Agostinho e Pitágoras com quem convivo todos os dias e viveram comigo cada momento dessa jornada e abriram horizontes.

Aos integrantes da banca pela boa vontade, dedicação e participação na avaliação desse trabalho.

À Faculdade de Letras da Universidade Federal de Minas Gerais por disponibilizar toda a estrutura necessária à evolução da pesquisa e ensinamentos adquiridos ao longo dessa trajetória.

Agradeço a Deus, pelo dom da vida e pela oportunidade de estar concluindo mais este trabalho que contribui para a realização de um sonho.

“Applied Linguistics is using what we know about (a) language, (b) how it is learned, and (c) how it is used, in order to achieve some purpose or solve some problems in the real world.”

(Schmitt and Celce-Murcia 2002, p.1)

ABSTRACT

This work is a case study with the objective of integrating online corpus activities into an English for Academic Purpose (EAP) classroom to help learners improve lexico-grammatical aspects of their writing. This work draws on Johns (1994), Chang (2010), O'Keefe, McCarthy, Carter (2007), Boutlon (2012), Tartoni (2012) and Frankenberg-Garcia (2012), who presented the benefits of integrating a corpus approach in the second language (L2) classroom for improving writing skills. The corpus chosen for the activities presented in this research was the Corpus of Contemporary American English (COCA). The participants were undergraduate and graduate students from various courses at UFMG, such as Engineering, Computer Science, Biology, Economy, Design, taking an English for Academic Purposes course. The corpus-based activities were introduced as part of the tasks related to a specific academic written genre: Statement of Purpose (SoP). The data collection took four meetings, which involved collecting students' writing before and after their accessing COCA. The students' experiences, using the tools while writing, were registered through Google Form questionnaires. Students' writings were marked and edited by the students. The results show that 79.25% of the attempts were successful. In some cases, the students' strategy was avoidance (15.38%) or no changes were made (16.67%). Therefore, it seems that learners are likely to improve their writing by accessing online corpora and by discovering for themselves frequent lexico-grammatical patterns.

Keywords: English as a Second Language; Online Corpora; Academic Writing; Lexico-grammatical Patterns; Editing Tools.

RESUMO

Este trabalho é um estudo de caso com objetivo de integrar o uso de corpus a uma aula de Inglês como segunda língua para ajudar alunos a melhorarem aspectos léxico-gramaticais da escrita através de consultas em um corpus online. Baseia-se em Johns (1994), Chang (2010), O'Keefee, McCarthy, Carter (2007), Boutlon (2012), Tartoni (2012) e Frankenberg-Garcia (2012), que apresentaram os benefícios da integração do uso de corpus em sala de aula para melhorar habilidades de escrita na segunda língua (L2). O corpus escolhido para essa pesquisa foi o *Corpus of Contemporary American English* (COCA). Os participantes desse estudo foram graduandos e pós-graduandos de diversos cursos da UFMG, como engenharia, ciências da computação, biologia, economia, design, cursando a disciplina Inglês para Fins Acadêmicos (IFA). As atividades baseadas em corpus foram apresentadas como parte das tarefas relacionadas ao gênero acadêmico *Statement of Purpose* (SoP). Antes de começar as atividades baseadas em corpus, os alunos tiveram um contato geral com o gênero estudado. A coleta de dados foi feita em quatro encontros de uma hora e quarenta minutos, cada encontro, com atividades específicas que envolviam coletar os textos dos alunos antes e depois da consulta ao COCA. A experiência dos alunos, ao usarem as ferramentas enquanto escreviam, foi registrada através do uso de questionários do *Google Forms*. Alguns erros foram marcados e editados pelos alunos. Os resultados mostram que 79.25% das tentativas de correção tiveram sucesso. Em alguns casos, a estratégia de alguns alunos foi de evitar o erro (15,38%) ou nenhuma alteração foi feita (16,67%). Os resultados encontrados nesta pesquisa evidenciam que aprendizes podem melhorar sua escrita através de consultas no corpus e descobrir por eles mesmos padrões léxico-gramaticais frequentes.

Palavras-chave: Inglês como Segunda Língua; Escrita Acadêmica; Corpora Online; Padrões Léxico-gramaticais; Ferramentas de Edição.

LIST OF FIGURES

Figure 1 - COCA web page showing the icons of the tools.....	33
Figure 2 - COCA web page – concordance lines generated by the list resource – occurrences of the word <i>good</i>	34
Figure 3 - COCA web page – Chart results of the word <i>good</i> presenting normalized frequency by section	35
Figure 4 - COCA web page - <i>Collocates</i> resources and the way they can be explored	36
Figure 5 - COCA web page - List of adjectives that come with the word “good” in the resource <i>Collocates</i>	36
Figure 6 - COCA web page – Concordance line of "blame" from COCA"	37
Figure 7 - COCA web page – Chart results of the word "blame" presenting normalized frequency by section	38
Figure 8 - COCA web page – <i>Collocates</i> , occurrences of prepositions before the word “Monday”	38
Figure 9 - COCA web page – <i>Collocates</i> , how verbs appear following “look forward to”	39
Figure 10 - COCA web page – <i>KWIC</i> with the word “book”	43

LIST OF TABLES

Table 1 - Absolute and relative frequency of the marked errors	49
Table 2 - Absolute and relative frequency of the errors separately	50
Table 3 - Type of lexical error followed by the relative frequency of successful correction	51
Table 4 - Type of grammatical errors followed by the relative frequency of successful correction	52
Table 5 - Reformulation attempts of Lexical and Grammatical Errors	52
Table 6 - Examples of errors that could have been caused by the L1 interference	54
Table 7 - Example of errors that could have been caused by distraction	54
Table 8 - Example of error that seem to be overgeneralization	55
Table 9 - Example of errors of prepositions	55
Table 10 - Examples of errors of articles	55
Table 11 - Examples of agreement errors	56

LIST OF GRAPHS

Graph 1 – Use of tools to assist writing the Draft - first questionnaire	46
Graph 2 – Use of tools to assist writing the SoP - second questionnaire.....	47
Graph 3 - List of COCA tools used by the students for editing their texts.....	48
Graph 4 – Percentage of lexical errors in each category and how learners faced the editing task	50
Graph 5 – Percentage of grammatical errors in each category and how learners faced the editing task	51

LIST OF ABBREVIATIONS

CEFR	Common European Framework Reference of Language
CL	Corpus Linguistics
CA	Contrastive Analysis
CIA	Contrastive Interlanguage Analysis
COCA	Corpus of Contemporary American English
DDL	Data Driven Learning
ESL	English as Second Language
GloWbe	Corpus of Global Web Based-English
IFA	Inglês para Fins Acadêmicos
IsF	Inglês Sem Fronteiras
GloWbe	Corpus of Global Web Based-English
KWIC	Key Word In Context
L1	First Language
L2	Second Language
SLA	Second Language Acquisition
SoP	Statement of Purpose

SUMMARY

1. Introduction	16
1.1 Motivation	18
1.2 Goals and Objectives	18
1.2.1 General objective	19
1.2.2 Specific objectives	19
1.2.3 Research questions	19
1.3 Organization of the thesis	19
2. Literature Review	21
2.1 Language teaching and learning nowadays	21
2.1.2 Error Analysis, Contrastive Analysis and Contrastive Interlanguage Analysis	23
2.1.3 Academic Writing	25
2.2 Corpus Linguistics and Corpora in the classroom	26
2.2.1 Corpus Linguistics	26
2.2.2 Corpora in the classroom	27
3. Methodology	32
3.1 Pilot studies	32
3.1.1 Activities	37
3.2 The study	41
3.2.1 The participants	41
3.2.1 The data collection	41
4. Results and Discussion	46
4.1 Discussion of the students' errors	53
5. Conclusion	57
References	60
APPENDIX 1 – First questionnaire	68
APPENDIX 2 - Responses	72
APPENDIX 3 – SoP Drafts	77
APPENDIX 4 - Workshop	81

APPENDIX 5 – Activities	91
APPENDIX 6- Errors selected from the second pilot study	103
APPENDIX 7 – Second questionnaire	105
APPENDIX 8 – Edited texts	107
APPENDIX 9 – Responses of the Second questionnaire	115
APPENDIX 10 – Tables with lexical and grammatical errors.	117

1. Introduction

As the world has become increasingly globalized, the Internet has become more accessible to educational institutions, and it is used as a support tool for teachers and students to perform their activities in the classroom. “In this changing world, institutions must learn to adapt quickly to the new demands of a highly literate and technological society” (MATTOS, 2011, p.34)¹. The development of literacy must take into account a variety of texts and forms of texts associated with new media and digital technologies (HEBERLE et al., 2015). The new technologies brought the need for new skills, which Lankshear and Knobel (2003) called “new literacies”.

The “new literacies” topic is related to this research by being part of a new concept of teaching and learning in the digital era we are living in, changing traditions, bringing new ways of learning, new media and vehicles. According to Lemke (1998, p. 283), literacies are technologies and they give the keys to using other technologies, broader ones. One of these new possibilities to access texts in an online environment is the use of corpora². A corpus is a collection of texts, written or spoken, or a mixture of both, selected according to given criteria, which are stored in a computer for analysis (O'KEEFFE; MCCARTHY; CARTER 2007) and also used as real language source in the preparation of textbooks.

With the possibility of accessing online corpora for free, teachers have the opportunity to introduce to their students activities that involve corpus consultation so that they can check their questions, edit their papers and, above all, raise their awareness about language use. Corpus consultation can be understood as a new literacy, which is referred as Corpus Consultation Literacy by O'Sullivan (2007). O'Sullivan (2007, p. 272) outlines that “learning is a process of knowledge construction and not a matter of absorbing and reproducing knowledge”. In this study, Corpus consultation is situated in the field of process-oriented instruction approach, which is an “instruction that fosters learning and thinking skills and metacognitive knowledge that is needed in the construction of understanding and use of knowledge” (DE JONG VAN HOUT-WALTERS 1994, p. 8).

¹ My translation of: “Nesse mundo em constante mutação, as instituições precisam aprender a se adaptar rapidamente às novas exigências de uma sociedade altamente letrada e tecnologicada”.

² Corpora is the plural of corpus. The word “corpus” comes from Latin and it means “‘body’, a set of something, in our case, a set of samples of language use” (BERBER SARDINHA et al. 2012)

This work also uses corpus tools in a Data Driven Learning (DDL) approach (JOHNS, 1994). DDL involves giving students access to sufficient language data, so that they can discover and learn for themselves rather than being taught the language explicitly (JOHNS, 1991). Corpora, corpus-analytic tools and corpus evidence “have been increasingly used in English language teaching and learning for the last two decades” (CHENG 2010, p. 319) and can effectively be used for learning, since online corpora could stimulate students and expose them to authentic language. Activities in which students are exposed to authentic materials can be effective alternatives to challenge and enhance language learning in the XXI century.

The advances in the direct access to corpora by language teachers and learners have created the need for research about a number of pedagogic issues. These issues include the types of corpora to be consulted, “large or small, general or domain-specific, tagged or untagged” (CHAMBERS 2005, p. 111); the kinds of learning strategies to benefit from direct corpus consultation; and the means by which direct access to corpora can be integrated into the language learning context. In sum, online corpora allow teachers and students to have access to texts of written and spoken authentic material and show how language occurs in real context. These tools can be useful for teachers to create tasks and for students to edit papers, giving them autonomy to improve both written and spoken texts.

After this brief introduction, the following sections will present the motivation (1.1) and the goals and objectives (1.2) of this work.

1. 1 Motivation

Writing in the academic context is a difficult task for many students, especially for the ones that need to write in a second language. This dissertation was motivated by the possibility of helping learners solve some of their writing problems. The idea is to use corpus as a tool for teaching and learning English. Corpora can be used to develop independence in academic discourse, for instance, helping learners to edit written assignments. A corpus is a more reliable guide to what language is like than human intuition and it can be used to discover information about language that may not have been noticed through intuition alone. It provides an evidence-based approach to language teaching and represents a speaker's experience of language in some domain. When consulting a corpus, the learner becomes the researcher and have the opportunity to discover some aspects of the language. That opportunity could give the learner the sense of achievement and independency

Corpus Linguistics studies has been growing with the advent of technology. Corpora have been available for years and little attention has been payed to it. After not long ago, corpora are seen by many teachers as a useful tool, but it is necessary to be aware of the possibilities of use and what it can offer. According to Sinclair (2004)

To make good use of corpus resources a teacher needs a modest orientation to the routines involved in retrieving information from the corpus, and – most importantly – training and experience in how to evaluate that information. It is the second point that has caused much controversy, because a corpus is not a simple object, and it is just as easy to derive nonsensical conclusions from the evidence as insightful ones. (SINCLAIR, 2004, p. 2)

The use of a corpus as a consulting tool for writing could have effects on the teacher, who would become a facilitator and guide, and effects on the learner, who would be able to learn how to learn and discover patterns of language.

1. 2 Goals and Objectives

The present section presents the goals: general objective and specific objectives of this study.

1.2.1 General objective

This investigation integrates corpus-based activities into an English language classroom to help learners improve their writing through the discovering of lexico-grammatical patterns. This data driven learning approach proposal involves online corpus consultation.

1.2.2 Specific objectives

1. Investigate how familiar the students are with online corpus tools;
2. Identify the types of errors in the students' Statement of Purpose (SoP) draft;
3. Analyze if they were successfully edited;
4. Analyze which tools the students reported they have used to edit their SoPs.

1.2.3 Research questions

The research questions of the present study were:

1. Are the learners familiar with online corpus tools?
2. Are their errors mostly grammatical or lexical?
3. Are the learners able to successfully edit their errors?
4. Do the learners tend to use more online tools for editing?

1.3 Organization of the thesis

This master thesis is organized as follows:

- Chapter 2 shows the Literature Review, where it is discussed language teaching and learning nowadays; Corpus Linguistics and its pedagogical applications

- Chapter 3 covers the Methodology, which describes the pilot studies and this investigation step by step;
- Chapter 4 brings the Results and Discussions;
- Chapter 5 presents the Conclusion.

2. Literature Review

This chapter presents discussions about language teaching and learning nowadays followed by definition of Corpus Linguistics (CL), some studies that involve CL, especially, studies concerning CL in pedagogy to show how CL studies have been growing. This spreading of CL is related not only to teaching and learning, but also to translation and language investigation. Error analysis will be covered since learner errors are analyzed in this thesis and Academic Writing, which is the specific purpose of the course in which this study was conducted.

2.1 Language teaching and learning nowadays

According to Lewis (2000), “learners don’t learn what teachers teach” (Lewis, 2000, p. 11), teaching is different from learning, while teaching is “organized, linear and systematic”, learning is “complex and non-linear”. Lewis (2000) affirms that it is not possible to control the amount of time students take to learn something and the amount of things they are able to learn. He also quotes Diana Larsen-Freeman’s footnote in an article of Applied Linguistics that affirms “teaching does not cause learning”, which brings the questions: what causes learning? What can be done for better learning?

In the XXI century, the new generation of learners demands new ways of teaching. Students have access to a large amount of information through television, books and, especially, Internet. Globalization makes access to information very easy and teachers are not seen as the center of knowledge anymore. Students have become independent and guiders may be teachers' best role. “Learners are active agents of the learning process and not passive absorbers of pre-constructed knowledge.” (OTERO at al. 2011, p. 17). According to Wong et al. (2015), students may learn when and what they want, however they want; it could involve technology and be called “seamless learning”.

Seamless learning implies that a student can learn whenever they are curious in a variety of scenarios and that they can switch from one scenario to another easily and quickly using the personal device as a mediator. These

scenarios include learning individually, with another student, a small group, or a large online community, with possible involvement of teachers, mentors, parents, librarians, workplace professionals, and members of other supportive communities, face-to-face or at a distance in places such as classroom, campus, home, workplace, zoo, park, and outdoors. Seamless learning space refers to the collection of the various learning scenarios supported by one-to-one technology. Exploration and investigation in the seamless learning space provides a potential to extend formal learning time, usually limited to the classroom, into informal learning time, to embrace opportunities for out-of-school learning driven by the personal interests of students, which may involve interacting with an online learning community, visiting museums, participating in community projects, or other venues. (WONG *et al.* 2015, p.5)

Seamless learning is not the focus of this work, but learning through corpus consultation could be somehow related to it.

Another advance in teaching and learning is the *flipped classroom*, which is a *blended learning* strategy, with the objective of improving students' engagement. *Flipped classroom* is a way of reversing the traditional learning environment, giving the students the chance to prepare themselves for the class at home and to be engaged in interactive activities in the classroom. Therefore, the *blended learning* implies that students have classes that they can attend online and then practice activities in the classroom. According to Mazon (2016), this model of learning “involves investigation and learning strategies based on Information and Communication Technologies”.³ This present research does not involve *flipped classroom*, but the use of technology and investigation as a learning strategy.

Recently, this investigation can be performed through the access of online *corpora* by using mobile phones inside or outside classrooms. The use of this device has been defended by several researchers (e.g. Bouhnik; Deshen, 2014; Mwakapina *et al.*, 2016), who stated that its use can generate interaction among students and teachers through Facebook and WhatsApp groups, use of dictionaries or websites for research, including research on corpora. Mobile phones were not used in this research, but *corpora* can be accessed on a mobile phone in case there is no computer in the classroom.

The idea of learners researching and that their “discovery” of the language is more effective than teaching the rules written in a textbook is not something new. Johns (1986, 1988) described

³ My translation of: “Trata-se de uma aprendizagem baseada na investigação, com estratégias educacionais apoiadas pelas Tecnologias da Informação e Comunicação (TICs)”. (MAZON, 2016, p. 124)

a microcomputer-based approach to foreign language teaching and learning which provides a context in which the learner can develop strategies to learn how to learn. This author proposed Data Driven Learning (DDL), which with the advance of technology also changed the way teachers teach and students learn. With the DDL approach, Johns uses concordancers to help students discover language patterns, how they happen and what occurs around it, rather than using a textbook with already determined rules. After years of working with concordancers, Johns (1994) was able to observe that DDL helps learners to develop the ability to see patterns in the target language. This will be discussed in more details in the next section, which will also cover corpus studies in general.

2.1.2 Error Analysis, Contrastive Analysis and Contrastive Interlanguage Analysis

When talking about ways of learning and teaching, it is important to mention error analysis, contrastive analysis and contrastive interlanguage analysis, which are part of this study. Error analysis, according to Khansir (2012), aims to “describe how learning occurs by examining the learner’s output and this includes his/her correct and incorrect utterances” (Khansir 2012, p. 1). Error Analysis cannot be separated from contrastive analysis, which is according to Andrade (s.d) a comparison between two language systems: (L1) and (L2). Some applied linguists have used contrastive analysis to enlighten their discussion. XIE Fang (2017) outlines its hypothesis by saying that

Contrastive analysis hypothesis claimed that the principal barrier to second language acquisition is the interference of the first language system with the second language system and that a scientific, structural comparison of the two languages in question would enable people to predict and describe which are problems and which are not. (XIE Fang, 2017, p. 10)

Errors are part of the students’ learning process and their interlanguage is a system. Interlanguage “refers to the separateness of a L2 learners’ system that has a structurally intermediate status between the native and target language learners” (XIE; JIANG, 2017, p. 11):

The beginning stages of learning a second language are characterized by a good deal of interlingual transfer from the native language. In the early stages, the native language is the only linguistic system upon which the learner can draw. These kinds of errors can be found in all aspects of language learning. (XIE; JIANG, 2017, p. 11)

It is not possible to be totally sure that some errors are caused because of interlingual transfer, which is only one of the types of errors, but knowing it is part of the learning process, it is possible to affirm that interlanguage may be influenced by first language.

Granger (1998) mentions a new strand of CA focusing on the interlanguage, called Contrastive Interlanguage Analysis (CIA). It is different from the traditional CA, which compares two languages, aiming to identify the differences. Granger (2012) outlines that when researchers compare learner's data and native data, that is called Contrastive Interlanguage Analysis (CIA). CIA is a step beyond error analysis as it contrasts competent speakers of English and learners' use and preferences in specific linguistic contexts and genre and according to Kimimoto et al. (1992), helps interpret the interlanguage data. As explained by Dutra; Silero (2010), the CIA can help researchers and teachers with linguistic aspects that should be treated pedagogically, such as the use of verb phrases. CL and the tools and consequent learning benefits are not so familiar among most second language acquisition (SLA) linguists (Dutra; Gomide, 2015), who have focused their studies mainly on the learners' mental process. Although Granger's 1998 book shed a light in the area by bringing and encouraging many other publications involving CL and SLA, there is still room for a great deal of research.

The growing interest in empirical studies have given support to corpus linguists that study learner production either oral or written. As suggested by Altemberg (2002), it is possible to investigate a certain language feature in a bilingual corpus and check if there are evidences of transferring on the learner's writing from L1 to L2.

The next section presents Academic Writing briefly, as it is part of this case study.

2.1.3 Academic Writing

Academic writing is a type of writing taught in an English for Specific Purposes (ESP) class, and it is a specific writing used in academic context, as the name suggests. International students who aim to study abroad for undergraduate or graduate programs must learn the different genres that are part of the academic world, such as Abstract, Summary, Essay, which are not part of this study, and Statement of Purpose (SoP), which will be better explained in the methodology of this study because it was the genre with which this study happened.

A common writing class differs from academic writing because the second, besides demanding different maturity and focus, it is backed up by evidence. It demands different technics for the different genres, such as avoiding plagiarism by paraphrasing, summarizing, referencing. The student also needs to learn the specific vocabulary used in these types of writings. Some academic courses assess students through writing, making academic writing very important to master before, in order to get into University, and during their academic life.

While dealing with writing, the English teacher must teach vocabulary, expressions, structure of the text. According to Hyland (2004)

An important principle here is to relate structures to meanings, making language use a criteria for teaching materials. This introduces the idea that particular language forms perform certain communicative functions and that students can be taught the functions most relevant to their needs. Functions are the means for achieving the ends (or purposes) of writing. (HYLAND 2004, p. 6).

This case study takes place in a course of English for Academic purpose, therefore academic writing is the kind of writing dealt with for the data collection. However, that topic is not going to be discussed further.

The next section presents CL and CL and classroom.

2.2 Corpus Linguistics and Corpora in the classroom

This section presents brief definitions, examples and a background of Corpus Linguistics studies. It is followed by the presentation of some studies related to corpora in the classroom.

2.2.1 Corpus Linguistics

Corpus Linguistics (LC) is a field in the intersection of Linguistics and Applied Linguistics studies and became more popular in the 1990s with the advent of more accessible technology and computers with better storage capacity. According to Berber Sardinha et al. (2012, p. 7), a corpus is a set of texts (written or spoken), “in computer file format, collected and organized with aim of being samples of a language or a linguistic variety”⁴. *The Cambridge Handbook of English Corpus Linguistics* brings the following definition of CL:

Corpus linguistics is a research approach that facilitates empirical investigations of language variation and use, resulting in research findings that have much greater generalizability and validity than would otherwise be feasible. (BIBER; REPPEN, 2015, p.1)

Corpus studies have changed as technology arises and it provides a great range of research possibilities, since it registers human communication as it is, without any adaptation. It allows people to discover language use as several tools can be used for research. Since it does not have a theoretical educational basis (SARDINHA *et al.*, 2012), it can be combined with any language teaching approach. The DDL approach (JOHNS, 1988) is an example of the use of CL combined to language teaching for specific purposes, which is the case of this case study as well.

The next subsection will cover corpora in the classroom and the DDL approach will be better explained.

⁴ My translation of: “em formato de arquivo de computador, coletados e organizados com a finalidade de serem amostras de uma língua ou de uma variedade linguística.” (SARDINHA et al. 2012, p. 7)

2.2.2 Corpora in the classroom

The amount of research involving corpus shows the state of the art in CL and in Linguistics in general: how much there is to be learned about language and how language studies can contribute to language teaching. Some CL researchers have proposed that CL can be used in language classrooms to facilitate learning. Corpus tools can be used in conjunction with other tools such as dictionaries to give students support in their tasks (O'KEEFFE, MCCARTHY; CARTER, 2007). It is possible to find out in online corpora how expressions occur in real contexts, while in a dictionary students confirm meanings or translations of words. Granger (2015) defends the use of a bilingual corpus for teaching and learning.

A corpus can be a tool for learners and teachers. Students can correct their own production, as proposed by Gilmore (2009), or the teacher can create exercises that could help learners to identify common errors. The teacher can create corpus-based exercises to be presented before or after a reading or listening activity, having the new input in several different contexts focusing on vocabulary, grammar and even touch cultural topics. Sonbul & Schmitt (2009) affirm that corpus-based activities are efficient to help student remember lessons and provide them a sense of accomplishment and commitment in dealing with authentic material. Discussions show that CL studies will revolutionize grammar teaching by 1. monolithic descriptions of English grammar will be replaced by register-specific descriptions; 2. the teaching of grammar will become more integrated with the teaching of vocabulary and 3. emphasis will shift from structural accuracy to the appropriate conditions of use for alternative grammatical constructions. (CONRAD, 2000, p. 549).

Reppen (2010) proposes a vocabulary class, which involved the compilation of a small newspaper article corpus by the students. They presented a list of words to verify the most frequent words before performing a reading activity related to these articles. The author also proposes an activity in which students were asked to generate word frequency lists from academic texts and from transcribed speeches.

In Brazil, there have been many studies related to corpora and languages. Some studies use corpora for learners' consultation (c.f. DUTRA, P. B. 2002; VICENTINI, 2006), some analyze learners' corpora and corpus compilation (c.f. LÚCIO 2006; PINTO 2008; DUTRA; GOMIDE,

2015; DUTRA *et al.* 2016), some propose instruments for analyzing learner's corpora (c.f. BERBER SARDINHA e SHEPHERD 2008), and others develop and analyze corpus of professional language (c.f. BERBER SARDINHA 2000; PARDO 2004). Some studies investigate specific topics and analyze, for instance, the use of degree adverbs in the writing of Brazilian learners of English as a second language (ESL) (c.f. DE OLIVEIRA, 2015). There are also studies that defend the use of concordance lines through DDL approach (TARTONI 2012). These studies provide results that somehow contribute to our understanding of learning languages.

Data Driven Learning (DDL), defended by Tim Johns (1986), is, according to Dutra (2009), one of the most innovative proposals “for highlighting the importance of the development of the students’ discovery ability, which induces them to find patterns and regularities in the language” (DUTRA, 2009 p. 83)⁵. DDL is an approach used for learners to discover through data. In language learning context, it helps students discover language patterns through data provided by the teacher, who previously obtained them from a corpus. According to Berber Sardinha (2004), DDL is one of the most solid proposals for the use of corpus material. Johns (1994) says that DDL approach is innovative in two aspects: one technological and the other methodological. Regarding technological aspect, it is innovative because it came with the increase of storage capacity and processing speed of affordable computers and the availability of ready-made corpora. It is innovative in methodological aspects due to the fact that it changes the roles of teachers and students in classroom: teachers’ new role is to help learners develop strategies to discovery and learn how to learn; learners’ become researchers. DDL “gives direct access to the data, so the learner can take part in building up his/her own profiles of meaning and uses” (Johns 1994, p.297). Johns (1991) used Key Word In Context (KWIC) in his research, which is a tool on corpus for generating concordance lines. Concordance lines are lines arranged by sorting criteria, for example: after choosing the main word, lines with occurrences of that word will show, making it possible to analyze what comes together with that word, how it appear, what are the contexts, what words show next to it etc. KWIC is a concordancer and it allows the user to arrange the concordance lines by sorting criteria, which involves choosing a main word and choosing how many words to the right and to the left. The DDL approach also provides learners access to authentic language, showing how language is actually used, instead of basing a class in prescriptive

⁵ My translation for “por colocar em evidência a importância do desenvolvimento da habilidade de descoberta do aluno, que o induz a encontrar padrões e regularidades na língua” DUTRA 2009, p. 83

grammar books. Through DDL, it is possible to find concrete examples of structures, words, verbs, and prepositions.

Frankenberg-Garcia (2012) presented some corpus-based activity based on DDL. In one of the activities, an ESL student from Portugal, whose first language (L1) is Portuguese, sought the translation for the word *conta* from Portuguese to English. After consulting a bilingual dictionary without success, the student used COMPARA (FRANKENBERG-GARCIA; SANTOS, 2012), a Portuguese-English online parallel corpus, and identified which word was the adequate translation for *conta* in the context she had chosen. Besides the activity mentioned, Frankenberg-Garcia (2012) demonstrates ways of learning structures by using concordance lines and have students analyze and together come up with a pattern of the use of a given verb. Frankenberg-Garcia (2012) affirms that the use of corpus can help teachers and students in three stages: before, during and after an activity. In a writing activity, for example, corpus can help with new vocabulary, answering questions students may have before or at the moment of writing (during), such as how to use prepositions, collocations for a specific context. After writing, students have some time available to edit and correct possible mistakes.

Some other pedagogical proposals involving the use of CL in the classroom, besides DDL, include the Lexical Syllabus and the Lexical Approach, as highlighted by Berber Sardinha (2004). The Lexical Syllabus was developed by Dave Willis (1990), and according to Berber Sardinha (2004), it is a proposal that uses the CL's methodology and considers lexicon as an isolated word. Lexical Approach emphasizes the lexicon compound feature, which means the words that accompany the lexicon.

Tartoni (2012) developed a study involving CL and corpus consultation in the classroom with DDL approach, presenting the concordance line to 9th grade students, who worked with comic texts and biography. She divided the class in two groups: one was going to work on their text using only a dictionary and the other group was going to work with the concordance lines generated in an online corpus and printed out to the students to consult. Tartoni (2012) noticed that the students who had access to the concordance lines showed more motivation when working on their texts, taking more risks, acquiring independency. As outlined by O'Sullivan (2007), corpus consultation involves some cognitive skills, such as: predicting, noticing, thinking, observing, analyzing, making inferences, which helps students acquire independency, corroborating with Tartoni's (2012) results.

Boulton (2010) presents a survey of 27 DDL empirical studies which focus on L2 outcomes. The countries and L1 involved were Malaysia (one study, Malay as L1), Oman (one study, Arabic as L1), Poland (one study, Polish as L1), Canada (two studies: one study had mixed L1 and the other had English as L1), Spain (two studies, one had English as L1 and the other had Spanish as L1), Taiwan (ten studies from which one had L1 marked with “?” and the others had Chinese as L1), Italy (one study, Italian as L1), Ireland (one study, mixed L1), Iran (one study, Farsi as L1), Germany (one study, German as L1), Sweden (one study, Swedish as L1), France (three studies, French as L1), the United States of America (two studies, one had mixed L1 and the other had English as L1). In most of the countries, the study has English as L2, except for one study from Canada, in which L2 was French; one study from the United States of America, in which the L2 was German; and one study from Taiwan, in which the L1 was marked by Boulton (2010) with a question mark (?) and the L2 was Chinese. The studies were performed in different ways, with different duration (hours, weeks or semesters), different amount of participants with distinct levels of proficiency (involving from low intermediate to advanced), some participants used a hands-on concordancer, a software including corpus data, or paper-based material. Boulton (2010) affirms that some researchers evaluate people’s affective reactions to corpus use through questionnaires asking them whether they found it easy, enjoyable, useful and the results were favorable, the majority of teachers and learners responded positively. The same was done in this research, in order to find out students’ opinions about corpus use. The aims were to check students’ outcomes of corpus use: it examines the outcomes of using corpus as a reference tool for writing, error correction, which is what this case study is about, or translation. The results are encouraging presenting objective and quantitative data from written performance evaluations.

Boulton (2008) tests DDL approach with lower level learners and describes a simple experiment on phrasal verbs with 113 lower-intermediate learners of English in their first year of university studies, with the following research question: “Can lower-level learners extract benefit from the raw data of a concordance print-out?” (BOULTON, 2008, p. 1). He exposed the learners to print-out concordance lines of authentic language, taken from the British National Corpus (BNC), to see if they could detect patterns of phrasal verbs use and apply their findings to new context. Boulton (2008) had examples of the base form of the verb “pick” and “picked” and the phrasal verbs “pick up” and “picked up”. He gave the students a pre-test in which the students had to select the base form or the phrasal verb and then gave them the concordance line print-outs.

After few minutes checking the occurrences of the phrasal verbs, the students had to do a post-test in which they had to apply the phrasal verb use to a new context. The results were encouraging “suggesting that even lower level learners can derive some benefit from corpus data, and that DDL should be added to their armory of techniques” (BOULTON, 2008, p. 1).

The next section will describe the methodology used in this case study, as well as the pilot studies.

3. Methodology

This chapter describes two pilot studies and the case study presented in this master's thesis. The first pilot study was conducted in order to find a suitable methodology for this research. This study was the format of a workshop with activities presented to students in order to teach them how to use online corpus tools.

The second pilot study was an attempt to collect the data, which did not achieve the expected objectives for the research. Afterwards, the steps for the case study data collection are shown, which includes the description of the students that took part in the study and the procedures to reach the results.

3.1 Pilot studies

In order to test the integration of corpora to English language classrooms, a pilot study was conducted. A workshop concerning the use of online corpora was presented to students from intermediate and advanced levels (B2) who participated in the program “English without Borders” (*Inglês sem Fronteiras - IsF*)⁶. Although this study does not use learners' corpora, it is important to emphasize that there is Corpus from IsF (CorIsF) being compiled mainly at UFMG (DUTRA; GOMIDE, 2015; DUTRA et al. 2016) with the collaboration of other universities, such as UFPA⁷ and UFPI⁸.

The way activities were used in the workshop for the pilot study and the study were inspired on Frankenberg-Garcia (2012) method, which involved giving students tasks in which they would have to use corpora to identify patterns. The activities on the Appendix 4, slides 4, 11, 12, 14, 15 and 16 were developed by Andressa Gomide together with the other teachers from *Núcleo de*

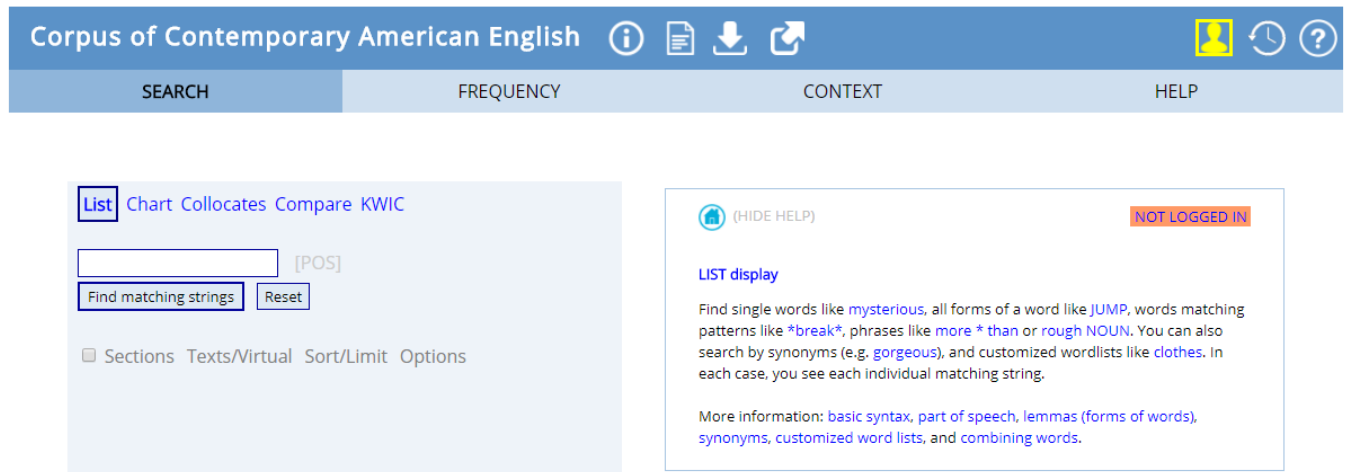
⁶ This program was launched in 2012 (DOU number 244, 12/19/2012, Section 1, pages 28/29.) In 2014, it became *Idiomas sem Fronteiras*, including the offering of other languages, such as French, German, Italian, Spanish and Portuguese as an additional language, impacting on the higher education process in Brazil. For details about the program, check Sarmiento et al. 2016.

⁷ Universidade Federal de Lavras

⁸ Universidade Federal do Piauí

Linguas (NucLi) at UFMG. The other activities were developed for the purpose of this study, in order to have students learn how to use COCA's different tools and explore possibilities of use. The tools available on COCA are List, Chart, Collocates, Compare and KWIC. Figure 1 shows the COCA web page with the icons of the tools. On the section "search", it is possible to type the words that are going to be searched and the results are shown on the "frequency" section. On "context", the corpus allows the user to have access to the original text or part of it.

Figure 1 - COCA web page showing the icons of the tools

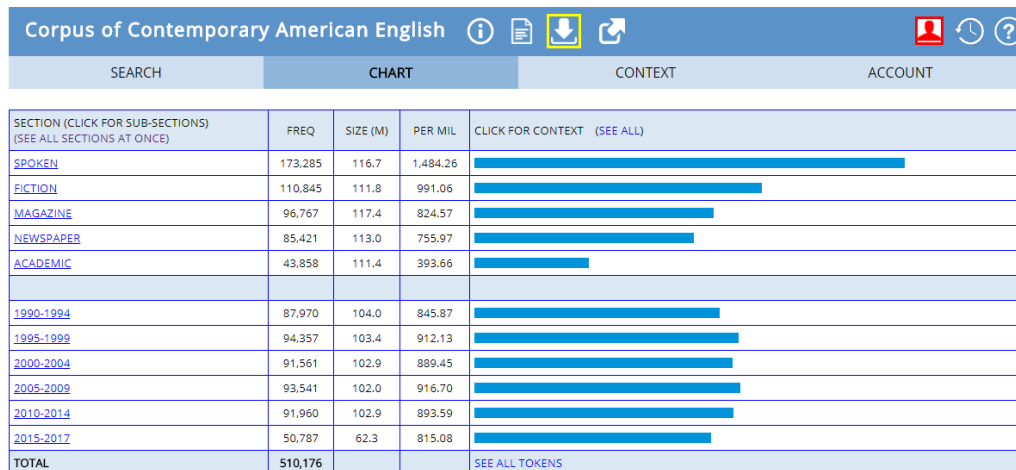


Three tools from the Corpus of Contemporary American English (COCA)⁹ were explored: *List*, *Chart* and *Collocates*. The tool *Chart* from the Corpus of Global Web-Based English (GloWbE)¹⁰ was also explored. Students were guided to learn how to use COCA. Figure 1 shows the screen with the *List* tool chosen

⁹ "The Corpus of Contemporary American English (COCA) is the largest freely-available corpus of English, and the only large and balanced corpus of American English. COCA is probably the most widely-used corpus of English, (...) which offer unparalleled insight into variation in English. The corpus contains more than 560 million words of text (20 million words each year 1990-2017) and it is equally divided among spoken, fiction, popular magazines, newspapers, and academic texts." (Available at <https://corpus.byu.edu/coca/>)

¹⁰ "The corpus of Global Web-based English (GloWbE; pronounced "globe") is unique in the way that it allows you to carry out comparisons between different varieties of English. GloWbE is related to many other corpora of English that we have created, which offer unparalleled insight into variation in English. GloWbE contains about 1.9 billion words of text from twenty different countries. This makes it about 100 times as large as other corpora like the International Corpus of English, and it allows for many types of searches that would not be possible otherwise. In addition to this online interface, you can also download full-text data from the corpus." (Available at <https://corpus.byu.edu/glowbe/>)

Figure 3 - COCA web page – Chart results of the word *good* presenting normalized frequency by section



The *Collocates* tool allows a user to verify the collocation of words searched. Collocation is, according to Sinclair (1991), “the occurrence of two or more words within a short space of each other in a text” (SINCLAIR 1991:170). Even though this study follows Sinclair’s (1991) definition, there are more definitions of collocation, such as Leheka (2015)

The concept of collocation is based on the notion that each word in a language prefers certain lexical contexts over others, i.e. that any given word tends to co-occur with certain words more often than it does with others. For example, the word *grass* is often used together with *green*, and the lexeme *LETTER* is often used together with the lexemes *WRITE AND READ* (LEHEKA, 2015).

It is possible to check items (nouns, prepositions, verbs, adjectives, adverbs, articles, conjunctions) that occur around other words, to the left or to the right (see Figure 4), in order to discover the words’ meaning and usage. For example, if you type “good” in the “collocate” display, check “noun” and choose 1 to the right, you will have a list of the most frequent nouns that occur right after the word “good”, which is shown in Figure 5. By clicking on them, it is possible to see sentences where the two words occur together, making it possible to check their meaning and usage.

Figure 4 - COCA web page - *Collocates* resources and the way they can be explored

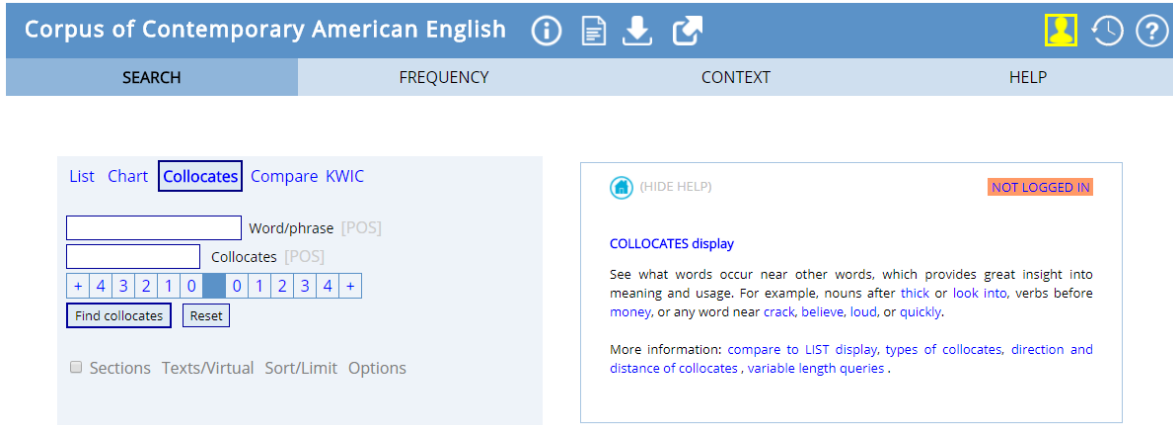


Figure 5 - COCA web page - List of adjectives that come with the word “good” in the resource *Collocates*

	<input type="checkbox"/>	CONTEXT	FREQ	
1	<input type="checkbox"/>	MORNING	21805	<div style="width: 100%;"></div>
2	<input type="checkbox"/>	NEWS	10363	<div style="width: 47%;"></div>
3	<input type="checkbox"/>	THING	9206	<div style="width: 42%;"></div>
4	<input type="checkbox"/>	IDEA	8348	<div style="width: 38%;"></div>
5	<input type="checkbox"/>	NIGHT	8112	<div style="width: 36%;"></div>
6	<input type="checkbox"/>	JOB	6973	<div style="width: 32%;"></div>
7	<input type="checkbox"/>	EVENING	6204	<div style="width: 28%;"></div>
8	<input type="checkbox"/>	TIME	6165	<div style="width: 27%;"></div>
9	<input type="checkbox"/>	LUCK	5619	<div style="width: 24%;"></div>
10	<input type="checkbox"/>	REASON	4180	<div style="width: 18%;"></div>
11	<input type="checkbox"/>	THINGS	3094	<div style="width: 13%;"></div>
12	<input type="checkbox"/>	FRIEND	2756	<div style="width: 12%;"></div>
13	<input type="checkbox"/>	QUESTION	2664	<div style="width: 11%;"></div>
14	<input type="checkbox"/>	WAY	2511	<div style="width: 11%;"></div>
15	<input type="checkbox"/>	PEOPLE	2307	<div style="width: 10%;"></div>
16	<input type="checkbox"/>	CHANCE	2200	<div style="width: 9%;"></div>
17	<input type="checkbox"/>	PLACE	2191	<div style="width: 9%;"></div>
18	<input type="checkbox"/>	DAY	2174	<div style="width: 9%;"></div>

3.1.1 Activities

The first activity proposed was to use the resource *List* and type the word “blame” (Figure 6). The students were asked to make a list of sentences where “blame” is a verb and a list of sentences where "blame" appears as a noun.

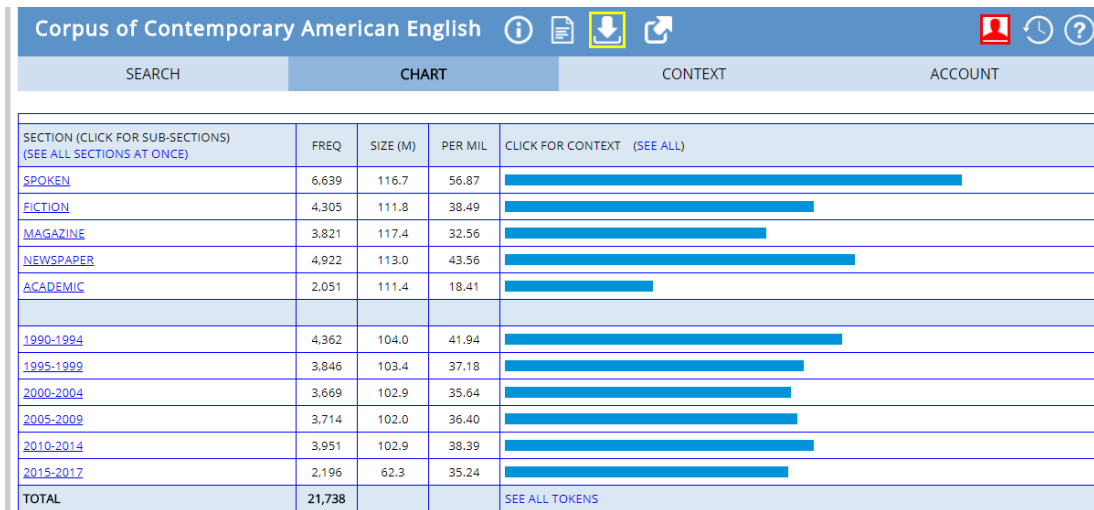
Figure 6 - COCA web page – Concordance line of "blame" from COCA "

The screenshot shows the COCA web interface with the 'CONTEXT' tab selected. The concordance list for 'blame' is as follows:

Line	Year	Genre	Source	Context
1	2017	NEWS	Chicago Tribune	A B C the truth. They delayed this, did their own investigation and then want to blame the FBI for not acting on it for 13 months.... "
2	2017	NEWS	Virginian-Pilot	A B C 4180366 # The blame lies in significant part with Oregon's out-of-touch elected prosecutors. These powerful forces within
3	2017	NEWS	Washington Times	A B C 4192555 # House Minority Leader Nancy Pelosi has a theory on who should shoulder the blame for over \$9 trillion in debt the nation has ar
4	2017	NEWS	Charlotte Observer	A B C storms began to pass through the area but officials are unsure if lightning is to blame.
5	2017	NEWS	New York Post	A B C This national-anthem ordeal has been out of control, and there's a lot of blame on myself, " Villanueva said Monday. " Obviously, we butcher
6	2017	NEWS	USA TODAY	A B C inspire them. # In conservative countries like Afghanistan and Egypt, some religious scholars blame the rising divorce rate on these shows. V
7	2017	NEWS	The Seattle Times	A B C didn't intend for the top managers to get bonuses. # He pointed the blame at then-CEO Ted Fick, who has since resigned under a cloud of cc
8	2017	NEWS	Detroit Free Press	A B C thank God some of my friends are. " # Dakich said he doesn't blame Izzo for his response, but he added that he believes that the Hall of
9	2017	NEWS	Colorado Springs Gazette	A B C unemployment is rising and the budget deficit has expanded. Many in the business community blame the slowdown on Bachelet's reforms c
10	2017	NEWS	Virginian-Pilot	A B C a few new ones. # The guild meetings are on hiatus for now -- blame busy schedules -- but several of those in the group are continuing their
11	2017	NEWS	New York Times	A B C come by. It's not an art form that forgives weakness, concession or blame. It runs on burnishing and destroying egos. It thrives on the perfor
12	2017	NEWS	Arizona Daily Star	A B C that an individual may have chosen homosexual behavior in good conscience relieves him from moral blame. # By contrast for Kelly and his
13	2017	NEWS	Arizona Daily Star	A B C 4180619 # According to Fiorina, much of the blame goes to good men and women who've chosen to remain silent when they see
14	2017	NEWS	USA TODAY	A B C last thing America needs is another loud voice angrily and unproductively telling you who to blame and who not to trust, " Abele said in a sta
15	2017	NEWS	New York Post	A B C may serve to remind voters of his pre-K triumph while allowing the mayor to later blame state and federal officials should funding dry up. #
16	2017	NEWS	Los Angeles Times	A B C White nationalists have spread the clip online via Twitter, using the hashtag #BLMkidnapping to blame the Black Lives Matter movement for
17	2017	NEWS	Chicago Tribune	A B C 's shooting drew withering criticism from Ferguson's office. The inspector general placed blame for the detectives' allegedly false narratives

Using the resource *Chart*, the students investigated in what register the word "blame" is more frequent (Figure 7).

Figure 7 - COCA web page – Chart results of the word "blame" presenting normalized frequency by section



Another exercise proposed was to use the resource *Collocates*, which allows the investigation of, for example, prepositions that come with days of the week (Figure 8) and months. Additionally, verbs that appear after expressions such as “look forward” were also explored, with the objective of observing how the verbs are followed by the morpheme -ING (Figure 9).

Figure 8 - COCA web page – *Collocates*, occurrences of prepositions before the word “Monday”



Figure 9 - COCA web page – Collocates, how verbs appear following “look forward to”

	CONTEXT	FREQ	
1	WORKING	1	
2	TOTEACHING	1	
3	SPEAKING	1	
4	SHOWN	1	
5	RELY	1	
6	IS	1	
7	HISTORICIZING	1	
8	GOING	1	
9	CAN	1	
10	ARE	1	
	TOTAL	10	

Students showed interest in using the online corpus and its tools. After noticing their reactions and receiving feedback from their teachers, who observed that the students were using the tool in class during writing activities, a Google form was created to assess students after a writing activity given by the teacher. The form presented questions about the tools, which were used spontaneously by the students while they did a writing task, which included an editing phase. The answers showed that they used the corpus, together with monolingual dictionaries and Google Translator. Thus, this pilot study involved learners of English and the use of corpus to give them opportunities to learn in which moments and how corpus tools can be used to improve their results in writing assignments.

A second pilot study was conducted in a class from the *Inglês para Fins Acadêmicos* (IFA)¹¹ course, which has a similar type of students. IFA is offered as a regular subject and attendance is more homogenous. The students are lower intermediate (B1) at an IFA class. The first moment consisted of having students write a first version of a text. After attending a workshop about the use COCA, they were asked to improve that version using COCA's tools. The errors in

¹¹ IFA is a course offered by FALE - UFMG to students from all fields that attend this university and require attendance and a sufficient grade for approval. This way, it is less likely that students skip classes or drop out. Similar to CorIsF, there is also another learner corpus being compiled at UFMG called *Corpus do Inglês para Fins Acadêmicos* (CoIFA). There are studies about CoIFA and CorIsF regarding students use of additive and result linking adverbials (DUTRA et al. 2017; DUTRA et al. in preparation). It is a great source of data since it is composed of different genre of texts produced by learners, such as summaries, statement of purpose, essays, literature review and articles.

the first version were not marked neither by the researcher nor by the class Professor¹². Even though students reported using the COCA tools, most errors were not corrected in the second version of their text. The conclusion was that due to their level of proficiency, they were not independent enough to look for their possible errors or to, at least reflect about them and check them in the online corpus. This data collection was then treated as a second pilot study. There was clearly a problem related to linguistic awareness. The students are low intermediate level (B1), and according to the Common European Framework of Reference for Languages (CEFR), B1 students

Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. Can deal with most situations likely to arise whilst travelling in an area where the language is spoken. Can produce simple connected text on topics which are familiar or of personal interest. Can describe experiences and events, dreams, hopes and ambitions and briefly give reasons and explanations for opinions and plans. (CEFR, 2001, p.29)

When it comes to writing, the CEFR states that B1 students “can write simple connected texts on topics which are familiar or of personal interest”, “can write personal letters describing experiences and impressions” (CEFR, p. 26). When it comes to qualitative aspects of spoken language use, regarding fluency, B1 students “can keep going comprehensibly, even though pausing for grammatical and lexical planning and repair is very evident, especially in longer stretches of free production” (CEFR, p. 7), which can explain their lack of independency when looking for their possible errors and not being able to reflect about them.

When similar activities were proposed to the *IsF* students, on the first pilot study, they were able to recognize and/or reflect about their own writing and the errors to be corrected. That could have been because they were higher intermediate (B2), which could explain the autonomy when editing their texts. According to the CEFR (p. 114), B2 students “show a relatively high degree of grammatical control” and they “can correct most of his/her mistakes” (CEFR p. 114). Due to these reasons, it was necessary to formulate a new strategy for the student to remember the class lessons involving another IFA class for data collection, which is described in the following section.

¹² This IFA class was taught by a Professor from UFMG and not a teacher.

3.2 The study

This section presents the steps for the data collection, as well as the participants and the activities. It also reports the steps taken for the data analysis, the codes used for each evaluated text and errors.

3.2.1 The participants

The participants were a group of twenty-two intermediate level (B1) students from IFA I, undergraduate and graduate students at UFMG from several courses, such as Engineering, Physiotherapy, Arts, Biochemistry, Veterinary Medicine, Computer Science, Anthropology, Biology, Fashion Design, Languages and Economy. Most of the students are Brazilian, except one, who is from Ecuador and is in Brazil pursuing his PhD in Arts.

The group was chosen based on the availability of the class professor, who, besides being involved in corpus research, was willing to allow this study to happen. Each IFA level focuses on a different academic written genre; IFA I studies Statement of Purpose (SoP), which is a one to two-page essay, for instance, “required as part of graduate student applications in the United States for master’s and doctoral programs” (SWALES 2011, p. 9). In a SoP, students should write about career trajectories, past achievements, distinctive individual qualities, personal and professional goals, reasons for choosing a particular specialization or location. It is also a genre that is part of the application process for mobility programs. At UFMG, the application for Minas Mundi¹³, UFMG major mobility program, includes the submission of a statement of purpose (*carta de intenções*) written in the language of the country the student wants to go to.

3.2.1 The data collection

The online corpus chosen for this research was COCA, because it covers a wide range of registers that may reflect the interests and specialization of the students, such as academic English

¹³ For more details access <https://www.ufmg.br/dri/wp-content/uploads/2018/04/Edital-Minas-Mundi-rev.pdf>

texts section, with subsections that covers History, Education, Geography and Social Sciences, Law and Political Sciences, Humanities, Medicine, Music. COCA is also easy to be accessed and, at the time of the data collection, UFMG contributed financially for the students to have access to it.

The data collection took four meetings of one hour and forty minutes. The students had previously attended a class in which their professor presented the general objectives of a SoP.

The activities and steps which were presented during the four days are described below:

- First day: The first part of the class consisted of a review of what a SoP is. Then, students were asked to share information with each other about the university they wanted to go to and why. The second part of the class was used to have them write a draft of their SoP within fifty minutes, which was a diagnostic written assignment. The students' drafts are shown on Appendix 3. Before the writing, they were asked to answer a short online questionnaire containing a consent term to use their text for research. The writing was done on Google forms and it was followed by questions about the tools they used while writing the text. The questionnaire is on Appendix 1 and the answers on Appendix 2.

- Second day: A workshop on how to use COCA was presented. It took one hour and forty minutes of activities, covering vocabulary and grammar tasks. The workshop covered the use of the tools *List*, *Chart*, *Collocates*, *Compare*, which were mentioned before on the pilot study, and this time KWIC¹⁴ (Figure 10) was added to the tasks. KWIC is a concordancer tool that produces lists of text segments in which words that co-occur with the keyword (in the center position) are arranged around it. The different colors indicate different parts of speech, making it easier to understand and analyze linguistic patterns. It was also presented to them the tool *Chart* on the Corpus of Global Web-based English (GloWbE), and Linguee¹⁵ as additional information.

¹⁴ KWIC (Keyword in Context) shows how a word can be used, how it appears in a sentence. Each part of speech has a different color.

¹⁵ Multilingual dictionary and search engine with access to a billion translations made by other people. Languages: Portuguese, English, Spanish, French and German. (webpage <https://www.linguee.com.br/>)

Figure 10 - COCA web page – KWIC with the word “book”

The screenshot displays the COCA web interface with the search results for the word "book". The interface includes a search bar at the top, navigation tabs (SEARCH, FREQUENCY, CONTEXT, ACCOUNT), and a list of search results. Each result shows a snippet of text with the word "book" highlighted in blue and its surrounding context in other colors. The results are numbered 1 through 28.

Rank	Year	Source	Context
1	1999	ACAD	PSAJournal
2	1991	SPOK	CNN_King
3	2002	ACAD	SocialStudies
4	2005	NEWS	Chicago
5	2012	SPOK	NPR_FreshAir
6	2006	SPOK	Fox_Sunday
7	2008	NEWS	AssocPress
8	2006	FIC	SouthernRev
9	1992	NEWS	Chicago
10	1996	NEWS	Atlanta
11	2012	SPOK	CNN
12	1996	FIC	BksDeepEndOcean
13	2009	ACAD	Writer
14	2006	NEWS	Atlanta
15	2006	NEWS	Chicago
16	2002	MAG	Essence
17	2004	ACAD	InfoSystems
18	1998	NEWS	WashPost
19	1994	SPOK	CBS_EyeToEye
20	1999	SPOK	CBS_SatMorn
21	1998	NEWS	Denver
22	2012	ACAD	Ref&UserServQ
23	2008	MAG	USAToday
24	1997	ACAD	NaturalHist
25	1993	SPOK	PBS_Newshour
26	2007	ACAD	Mercury
27	2006	NEWS	Chicago
28	1994	NEWS	CSMonitor

On the first activity, the students were asked to investigate the word “blame” using the tool *List* and separate the occurrences in which “blame” appears as a verb and occurrences in which it appears as a noun. After that, they were encouraged to check “blame” on “KWIC” and see the colored words around it, observing and analyzing the function of each word. The second activity covered the use of the tool *Chart*, where the students were supposed to find out the use of the words “however” and “but” on different sections, which are “spoken”, “fiction”, “magazine”, “newspaper” and “academic”. The third step was to teach them how to use *Collocates* by asking them to find out if “homework” is used after “make” or “do”. After that, they were asked to find the prepositions that occur with some words shown on the slide presentation. Still using *Collocates*, students were asked to, in pairs, find verbs and prepositions that appears after or before some words given on the slide presentation. The fourth activity was given with the use of GloWbE corpus; the students had to choose *Chart* and analyze the occurrence of the words “color” and “colour” in the different sections, which were countries, such as The United States, Great Britain, Australia, Canada, Ireland, among others. The fifth activity was to find prefix of words on COCA by typing on the display of *List* the wildcard (*) before some words given. The last part was

teaching them how to find synonyms by typing the formula [=word], also generated on the tool *List*. The activities presented at the workshop are on Appendix 4.

- Third day: The class concerned learning how to edit errors by using an online corpus. The researcher analyzed the students diagnostic SoP to find out the types of errors they had made. There was CIA when the researcher compared the students' writings to competent speakers of English on the corpus. Students were shown examples of errors made by them on their writings and explanations (see Appendix 5). They were given activities (see Appendix 6) created by the researcher for them to, in pairs, correct similar errors to the ones they had on their texts. These errors were selected based on the possibilities to be solved and corrected by checking occurrences on COCA. The students were also asked to correct errors, previously marked, from other students' writings, who were at same level as they were¹⁶. After the correction attempts, the strategies they had chosen to find the correct form of what was marked were discussed. They came up with various possibilities that were shared with the class. Some solutions were effective, some were not.

On this stage of the collection, there was clearly "social learning" (O'SULLIVAN, 2007, p. 274), which is part of process-oriented instruction approach (DE JONG VAN HOUT-WALTERS 1994, p. 8), mentioned earlier in this work. According to O'Sullivan (2007), "social learning" occurs when students have an opportunity to learn from those around them in the classroom. The classroom can be considered a social learning environment. "Social learning" is defended by Vygotsky (1978) as a cognitive development and social interaction. The students had a chance to share the choices they made on corpus, how they found the solution for the problem and there was more than one way.

- Fourth day: Students were asked to edit and continue writing their SoP (those who had not been able to finish it). The editing was based on the researcher's feedback which was done on each student individual text by highlighting the errors (Appendix 3). The errors were marked (highlighted in yellow) considering the possibility to be checked on the COCA tools they had learned to use. In order to predict the students steps when searching on COCA, the errors were checked in advance by the researcher. Words, such as nouns, verbs, adjectives, nouns following

¹⁶ These writings were the students' texts from the second pilot study.

verbs and verbs followed by prepositions were marked. Students were asked to edit their texts individually and this time they were supposed to add COCA as a resource for editing/writing. A second questionnaire (Appendix 7) followed the writing assignment and students were asked about the tools they used and if they considered them useful for their writing process.

Research data were collected in two ways:

- Application of two questionnaires on Google Forms to poll the students' opinions about the tools and their writing process. The first questionnaire was given after they wrote their SoP draft (see Appendix 1), and the other after writing the second version (see Appendix 8).
- Analyses of the students' writings, lexical and grammatical errors. In order to make the analyses, it was necessary to create a code to mark the sentences to be edited. The code includes the text and the sentence numbers, for example, **T1-2** refers to Text 1 and sentence 2.

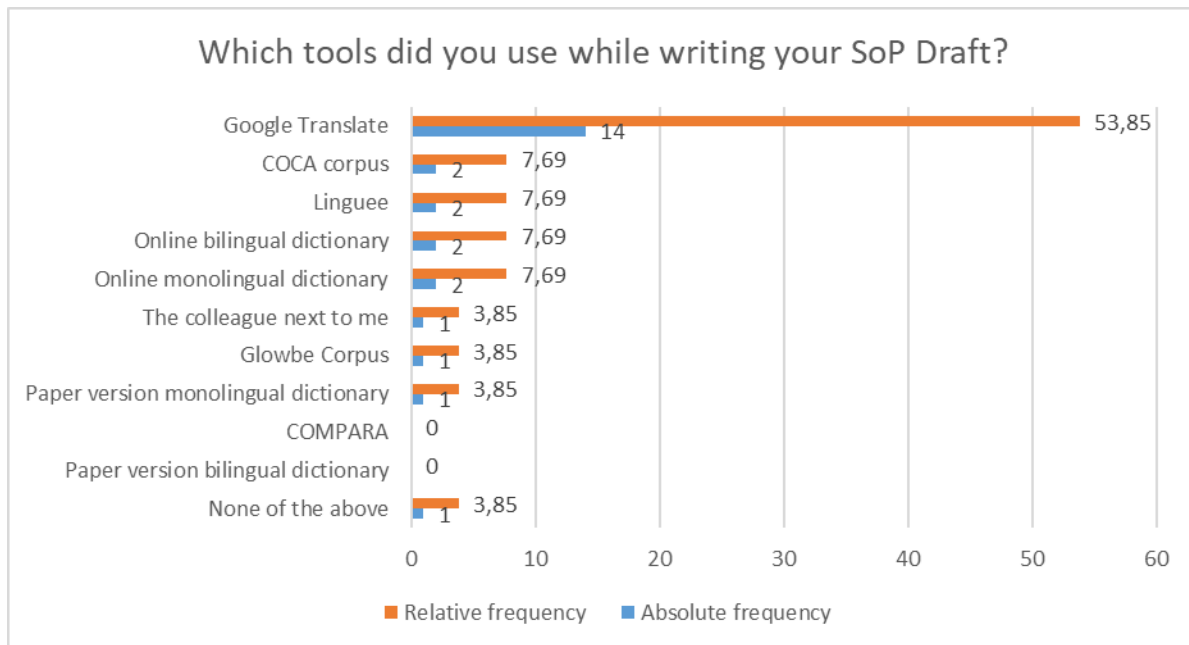
The following chapter presents the results and the discussion of the results.

4. Results and Discussion

This chapter presents the research results followed by the discussion of such results.

From the texts produced by twenty students, only thirteen texts were used and analyzed in this study due to lack of attendance of the other students in some of the data collection days. The students were asked to answer to 2 questionnaires: one after writing the SoP first draft (Appendix 1) and another one after writing the second version (Appendix 8). The answers on the first questionnaire (Graph 1) showed that, for the draft writing, 53.85% of them used Google Translate and few of them used monolingual, bilingual dictionaries and Linguee. There were 2 students who used COCA, one of them is an International Economics Relations student and the other one is a French major student.

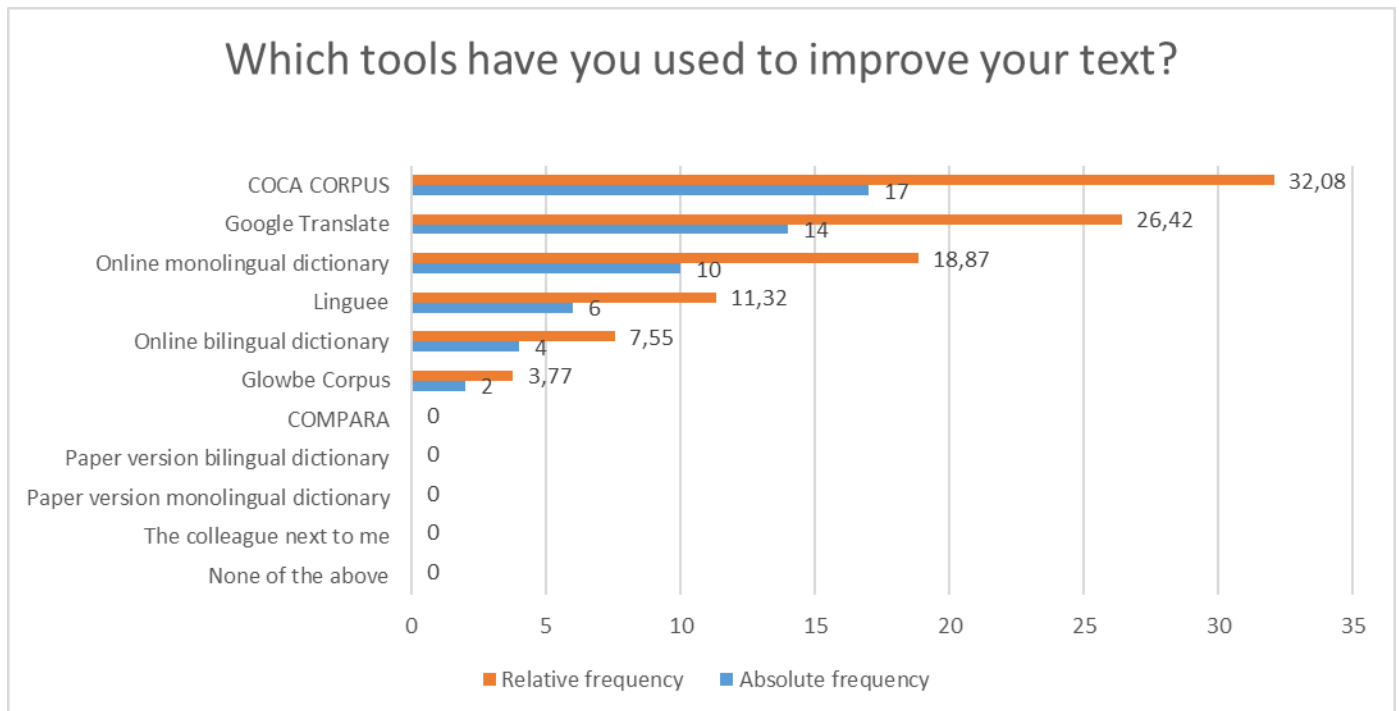
Graph 1 – Use of tools to assist writing the Draft- first questionnaire



At the time this group wrote the SoP diagnostic text, most of them relied on Google Translate as a writing tool. This tool can be used to translate individual words, groups of word, sentences or paragraphs. Yet, students did not report how they used Google Translate.

For the writing of the second version, students were supposed to edit the previous version. This time, they reported using more tools than they used when writing the first draft (Graph 2).

Graph 2 – Use of tools to assist writing the SoP- second questionnaire



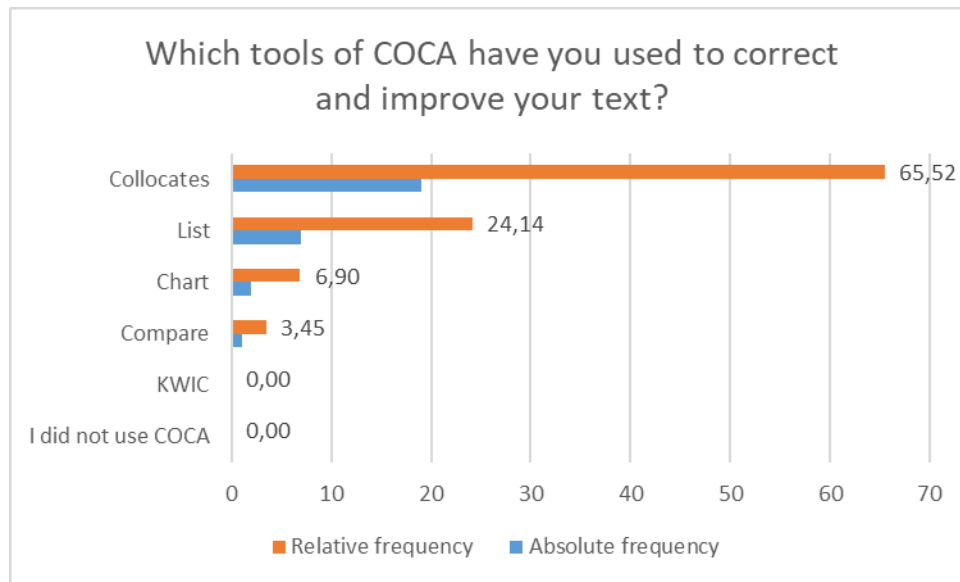
The highest increase in the use of the available tools was related to the use of corpus. This was expected since the workshop allowed the students to use COCA in two classes. However, there was also an increase in use of the other tools. This behavior seems to be because the students, once with the opportunity to edit their texts, might have needed to look for solutions, and, therefore, used different tools.

The editing moment was an individual task, which might have been the reason why “the colleague next to me” did not appear on Graph 2. Students also increased the use of COCA, GloWbE and Linguee. COCA and GloWbE are corpora, Linguee¹⁷ is a data bank of texts and they were presented at the workshop. Although, students were asked to use COCA when editing, the other two were not required to be used. Since students had the chance to learn how to use them,

¹⁷ Linguee cannot be considered a corpus; it is a data bank of aligned translated texts and their originals. On its website, Linguee is presented as a bilingual dictionary, but it fits a description of “translation corpora” given by Altenberg (2002) which is “contain texts in the original and translated versions” (ALTENBERG, 2002, p. 38)

such tools were seen as suitable to be used during the editing period of the class. The main objective was to investigate how the students used COCA; consequently, they were also asked to answer which of these corpus's tools they preferred to use during the editing phase. As shown on Graph 3, the tools they used the most was *Collocates*.

Graph 3 - List of COCA tools used by the students for editing their texts



The activities presented in the workshop (see Appendix 4) were varied; in other words, their intention was not to induce the students to use collocates more than the others. Besides having the opportunity to explore *Chart* and *KWIC*, they were also exposed to different ways of using *List*, such as finding prefixes of any word (by using the *wildcard* (*) and typing the word on the list display) and finding synonyms by typing [=word]. What might have caused the high use of collocates was the kind of errors market for editing.

The texts were analyzed and tables with the students' errors were created followed by the student's new version (see Appendix 10). The code "A" was used for the cases of reformulation of what was previously marked, being "A1" as reformulation with success and "A2" as reformulation without success. There were also cases in which the student did not reformulate what was previously marked and this was coded as "B" for no reformulation. Besides not reformulating, there were students who avoided the marked item by deleting it from the new version. Those cases were coded as "C" for avoidance. Table 1 shows the general data with the

absolute and relative frequency of the marked errors of each category of reformulation or non-reformulation.

Table 1 - Absolute and relative frequency of the marked errors

Type of error	Error description	Number of errors	Percentage
A1	Reformulation with success	42	53.85%
A2	Reformulation without success	11	14.10%
B	No reformulation	13	16.67%
C	Avoidance	12	15.38%
TOTAL		78	100%

The errors were divided into lexical and grammatical errors (Appendix 10), so that we could analyze the errors and investigate if there was a more successfully edited category. The errors marked were the ones that students could check on COCA if they decided to. There were other errors made by the students, such as related to coherence, conjunctions, punctuations, organization, but they were not the focus of this study. We could also check which type of error was more frequent among the students, considering also that the researcher was the one who decided which ones to mark. When marking the errors, there were no criteria for marking specific lexical and grammatical ones.

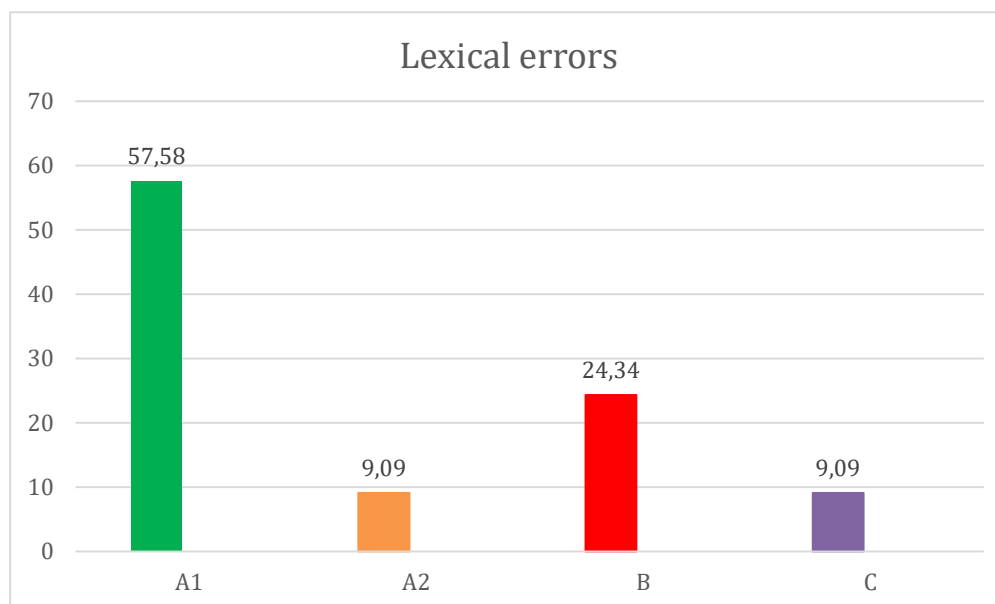
It was necessary to categorize the errors as lexical and grammatical errors due to observation of the kinds of errors and the amount of errors marked and corrected. Thus, the first analysis was done by categorizing the errors into grammatical and lexical errors. The errors were divided into two tables generated in an electronic sheet, one table with lexical errors and one table with grammatical errors (see Appendix 10) together with the amount of errors corrected and not corrected and marked with the codes shown in Table 1. Among the errors marked, 33 were lexical and 45 were grammatical. The frequency of the errors was converted into percentage and then presented on Table 2 and on bar graphs. Graph 4 shows the amount, in percentage, of lexical errors and Graph 5 grammatical errors.

Table 2 - Absolute and relative frequency of the errors separately

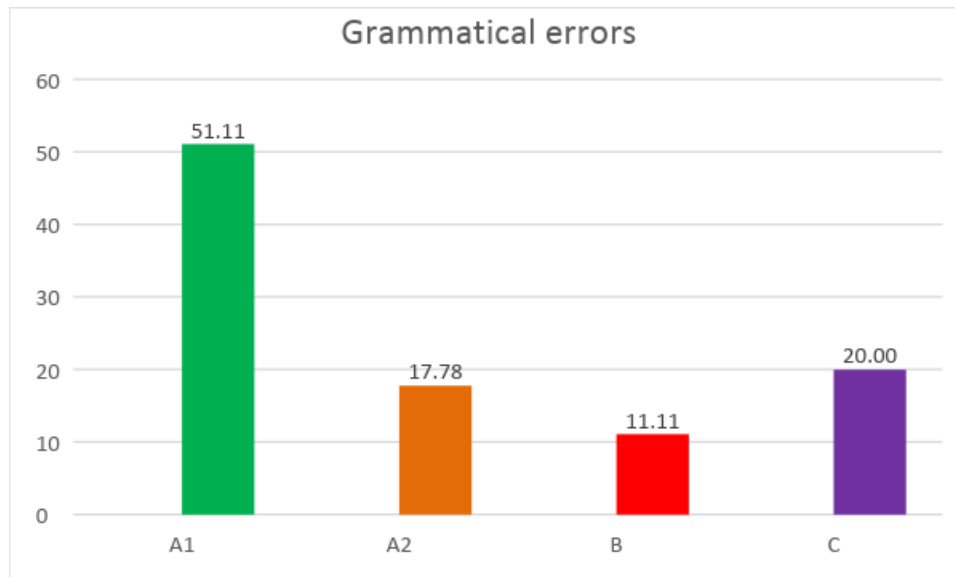
Type of error	Error description	Lexical error		Grammatical error	
		Absolute frequency	Relative frequency	Absolute frequency	Relative frequency
A1	Reformulation with success	19	57.58%	23	51.11%
A2	Reformulation without success	3	9.09%	8	17.78%
B	No reformulation	8	24.24%	5	11.11%
C	Avoidance	3	9.09%	9	20.00%
TOTAL		33	100%	45	100%

Table 2 presents the results after an analysis of all the texts together. There were 33 lexical errors, from which 19 were corrected with success, which represents 57.58%. There were 45 grammatical errors, from which 23 were corrected with success, which is 51.11%. Despite the cases in which there were avoidances or no changes, a small percentage failed when trying to correct, which are 9.09% lexical and 17.78% grammatical errors.

Graph 4 – percentage of lexical errors in each category and how learners faced the editing task



Graph 5 – Percentage of grammatical errors in each category and how learners faced the editing task



Among the lexical errors marked, some of them were probably due to similar pronunciation between the word they should have used and the word they used (homophone problem), others were words chosen wrongly and seemed to be due to first language (L1) interference on the second language (L2) like false cognates, a wrong translation from L1 to L2, students tend to use the first corresponding word without considering semantics. There were also errors that could not be explained by the researcher, such as words out of place in a sentence as in “My name is xxxx and a study english for **a any** years (sic)”. Most of the lexical errors seem to be caused by L1 interference which accounted for the highest number of successful corrections. Table 3 shows the relative frequency of the amount of successful correction among those who tried to correct.

Table 3 - Type of lexical error followed by the relative frequency of successful correction

Lexical error type	Relative frequency
L1 interference	81.81%

As for the grammatical marked errors, they were errors of structure, like subject verb agreement, lack or wrong prepositions, article errors, verb tense errors. Most of the grammatical errors were structure related and they were also the ones with more successful corrections. Table

4 shows the relative frequency of the amount of successful correction among those who tried to correct.

Table 4 - Type of grammatical errors followed by the relative frequency of successful correction

Grammar error type	Relative frequency
Structure	84.61 %
Preposition	33.33%
Article	100%
Tense	57.14%

Some of these errors could also have been caused by L1 interference.

Table 10 and Table 11 show the lexical errors and the grammatical errors generated (cf. Appendix 10). They show the lexical errors and grammatical errors followed by the categories of reformulation with success (A1), reformulation without success (A2), no reformulation (B) and avoidance (C).

Among the lexical errors, there were 22 correction attempts (66.67%) (group A), 19 (57,57%) were successful (A1). As for grammatical errors, there were 31 correction attempts (68.89 %) (group A), 23 (51,11%) were successful (A1). See Table 5.

Table 5 - Reformulation attempts of Lexical and Grammatical Errors

Type of error	Error description	Lexical error		Grammatical error		Total	
		Absolute frequency	Relative frequency	Absolute frequency	Relative frequency	Absolute frequency	Relative frequency
A1	Reformulation with success	19	86.36%	23	74.19%	42	79.25%
A2	Reformulation without success	3	13.63%	8	25.81%	11	20.75%
Total		22	100%	31	100%	53	100%

Since group A encompasses the errors that were attempted to be corrected by students, we can look at them separately. Out of the 22 lexical errors, 19 of them (86.36 %) were successfully corrected and only 3 (13.63%) were not. If we consider the grammatical errors, 23 out of 31

(74.19%) were edited correctly while 8 (25.81%) were not. The successful corrections of the lexical errors were higher than the successful corrections of the grammatical ones. In total, among 53 correction attempts, both lexical and grammatical, 42 (79.25%) of the attempts were successfully corrected and 11 (20.75%) failed. These results are important because they show the high motivation the students had for editing and improving their writings.

The other two categories proposed to analyze how the students faced the marked errors cannot be considered either a success or a failure. The C cases were not considered failure or success due to the fact that it does not apply to the use of the tool. The same applies to the B cases, since the tool was not used to make any change upon what had been written. The use of COCA tools led to effective editing, due to the fact that there were more than 50% successful corrections and a small number of attempt failed.

4.1 Discussion of the students' errors

In this section, the errors made by the students on their writings will be discussed according to the literature.

The lexical errors were homophone induced errors, false cognates, orthographic errors, wrong words that in L1 (in this case, Portuguese) can be used in many situations but in English it does not apply. The grammatical errors were mostly structure related to agreement errors, tense, singular-plural, prepositional errors, article and conjunctions errors.

In order to analyze both lexical and grammatical students' errors, it is necessary to consider their possible reasons. The majority of errors seem to have been caused by L1 system interference in L2 system. It is not possible to be 100% certain that the students' errors were because of L1 interference, which makes it is important to quote Xiao one more time when he says that "language acquisition occurs in the mind of the learner, which cannot be observed directly and must be studied from a psychological perspective" (XIAO 2007, p. 1), which was not done in this case study.

Some examples of the errors made by the students that could have been because of the L1 (Portuguese) influence, are presented at Table 6.

Table 6 - Examples of errors that could have been caused by the L1 interference

Sentence code	Sentence
T1-5	others researchers
T4-2	We public papers of this subject in Revista Trama (Rio The Janeiro)
T8-1	When I decided to course Letras, during my classes I started to being very excited about the studies in language acquisition.
T9-3	who I had the pleasure to live with and divide academic experiences.
T1-4	My research is about de middle of learning between observation and reproduce, indeed, my research is about “the copy”.

T1-5 used plural in both adjective and noun. T4-2 used the word “public” rather than “publish”, which is a false cognate. T8-1 used the word “course” as a verb, which is similar to the way that word appears in Portuguese. T9-3 used the word “divide”, which is the one used in Portuguese, instead of using the appropriate word in English (share). In T1-4 example the student probably checked the translation of the word “entre” in Portuguese to English and he/she had “between” as a result.

Some errors seemed to be due to distraction, a failure to utilize a known system correctly, more like a mistake, as it is shown in Table 7.

Table 7 - Example of errors that could have been caused by distraction

Sentence code	Sentence
T3-6	I work since twelve years, because in city when i borne, they accept all the kids
T9-6	now I have interest to know more about public politics and how to make a word a better place to live.
T8-3	I started to work in a project that was dedicated to knew professor , to be their first jobs and know about their area a little more.

On T3-6 the student wrote “when” instead of “where”; on T9-6, the student misspelled the word “world”. On T8-3, the student misspelled the word “new” because of the sound similarity,

which is an example of homophone, a word that is pronounced the same but differ in meaning. These kinds of errors are not the kinds that are going to reoccur because they seem to be a matter of distraction.

In Table 8, the error presented seems to be overgeneralization, which is a kind of interlanguage transfer according to XIE Fang, JIANG Xue-mei (2017). The student seems to know that the past tense of “take” is “took”, but still added the “ed” mixing the regular and irregular form of the past simple. That is the generalization of the use of “ed” to form the past.

Table 8 - Example of error that seem to be overgeneralization

Sentence code	Sentence
T8-4	last semester I toked class of Oral Expression that was a discipline that to help us with our English skills

The tables bellow show examples of some other errors, such as errors of prepositions (Table 9), errors of articles (Table 10) and errors of agreement (Table 11).

Table 9 - Example of errors of prepositions

Sentence code	Sentence
T9-7	My love of Economic started during my Post-Graduation Lato Senu Course on Public Law field
T1-3	I would lik to know try make a exchange with on university on USA or other english speaker country for improve my knowledge and my conversation.
T11-1	i hope to work on Material Science as research science or make my own business.

T9-7 used “of” instead of “for”, T1-3 used “on” instead of “in” and forgot the definite article “the”, T11-1 used “on” rather than “with”.

Table 10 - Examples of errors of articles

Sentence code	Sentence
T1-3	I would lik to know try make a exchange with on university on USA or other english speaker country for improve my knowledge and my conversation.
T11-3	First of all, I wish to get a undergraduate

T1-3 used an instead of “an”, T11-3 also used “a” instead of “an”, may be because once he/she learned that before “u” the indefinite article is “a” (in “University” for example), but did not considered the sound. This could also be considered an overgeneralization case.

Table 11 - Examples of agreement errors

Sentence code	Sentence
T3-2	I am loved Biology
T3-5	i am research about influence of MAS receptor

In the beginning of sentence T3-2 the first person, singular, appears followed by the verb “love” in the past, which is also an error of tense; T3-5 also fails to agree “I am”, which is present tense, first person, singular and did not conjugate the verb “research” properly to the person and the tense.

The final chapter presents the conclusion and final considerations, in which the research questions are answered and the conclusions of this work are presented.

5. Conclusion

This work is a case study with the objective of integrating online corpus activities into an English for Academic Purpose (EAP) classroom to help learners improve lexico-grammatical aspects of their writing through the discovering of language patterns. The specific objectives of this research were to investigate how familiar the students are with online corpus tools; identify the types of errors in the students' Statement of Purpose (SoP) drafts; analyze if they were successfully edited and analyze which tools the students reported they have used to edit their SoPs.

On the pilot studies, it was possible to test the steps for this research and how to do it. The second pilot study, which was at first an attempt to the study, brought the conclusion that due to the level of proficiency of the B1 students, they were not independent enough to look for their possible errors or to, at least reflect about them and check them in the online corpus. The errors on the students' writings had to be marked.

This case study was based on a DDL approach by exposing the learners to authentic material, providing independence and autonomy for correcting their errors and through that having access to patterns of the use of the language. O'Sullivan (2007) defends that the teacher should show how content is acquired by modelling "learning" and showing learners how to proceed with the learning process, which was efficiently done on the four meetings with the learners through the workshop and the activities with social learning.

It was possible to conclude that the students were not familiar with online corpus tools (except for two of them, who had taken courses at FALE before) and after being introduced to them, they could discover more about English, as well as develop their capacity of investigation and awareness of this language. Another conclusion is that *Collocates* was the most used COCA tool by the students when attempting to correct their own errors. The corpus consultation could help finding out the words that occur together with the word in question, therefore, how that word appears in a sentence.

When analyzing the lexical and grammatical errors, there were more grammatical errors marked than lexical. It was not because the students made more grammatical than lexical errors, but because the number of errors to be marked was not the focus, therefore a specific amount had

not been established by the researcher previously. The errors were marked based on the activities they were asked to do on the workshop and on the possibility for the students to find a solution for the errors on COCA. The lexical errors were homophone induced errors, false cognates, orthographic errors, wrong words that in L1 (in this case, Portuguese) can be used in many situations but in English it does not apply. The grammatical errors were mostly structure related, agreement errors, tense, singular-plural, prepositional errors, article and conjunctions errors.

The participants' reactions were favorable to the use of corpus, although their appreciation may be subjective. The use of corpora among students can be used as a reference resource, together with dictionaries, translators among others. Most students used more than one search tool when they were asked to correct their texts. This may be due to the fact that they were in the position of editing, which stimulates the necessity to look for more information.

The use of concordance lines in class seems to be beneficial to lower-level students as well as it is for higher-level ones, which corroborates Boulton (2010) and answers Johns' (1994) question about the possibility of DDL being applied to beginners or near beginners. The participants of this study were low intermediates (B1), and the results were favorable to the use of corpus consultation as well. They did not use concordance lines, as described in Boulton (2010) and Johns (1994), but reported the frequent use of collocates, which present the co-occurrence of words immediately before and after a keyword – not as word sequences as in concordance lines.

The use of COCA tools led to effective editing, since there were more than 50% successful corrections and a small number of attempt failed. This shows significant improvement on lexical and grammatical awareness. These results corroborate Tortoni's results (2012). She worked with 9th graders, used DDL approach presenting the concordance line and the majority of the students demonstrated to be encouraged and motivated to participate after having contact with the concordance lines in one of the activities proposed by her. It also corroborates with the results of Boulton's studies presented on his article, which show favorable results to affective reactions to corpus use, encouraging results to the editing and error correction results, and the effectiveness on B1 students. Finally, besides being useful for writing correction, it certifies what has been tested by Johns (1994) due to the fact that students were able to discover language through corpus research, an exercise that cooperates with learning and because of such discoveries that they were successful.

This was a case study and the results of the present study can only support assumptions made on the data found in the study. Further studies can be planned to make more accurate claims. More time might be necessary to investigate deeply the effectiveness of the use of corpora by students. As a possibility of future studies, it is possible to compare the SoP produced during this study to the ones which are already included at CorIFA. It could lead to results associated with error analysis research.

References

ALTENBERG, Bengt. Using bilingual corpus evidence in learner corpus research. **Computer learner corpora, second language acquisition and foreign language teaching**, v. 6, p. 27-54, 2002.

ANDRADE, Fabrícia. Análise de erros em produção escrita de estudantes de inglês como língua instrumental.

AŞIK, Asuman; VURAL, Arzu Şarlanoglu; AKPINAR, Kadriye Dilek. Lexical awareness and development through data driven learning: Attitudes and beliefs of EFL learners. **Journal of Education and Training Studies**, v. 4, n. 3, p. 87-96, 2015.

BERBER SARDINHA, A. P. *Linguística de corpus*. Barueri: Manole, 2004.

BERBER SARDINHA, A. P.; SHEPERD, T. An online system for error identification in Brazilian learner English. In: **Anais do 8th Teaching and Language Corpora Conference**. 2008. p. 257-262.

BERBER SARDINHA, T., SHEPHERD, T., DELEGÁ-LÚCIO, D., & FERREIRA, T. *Tecnologias & mídias no ensino de inglês—O corpus nas “receitas”*. 1 ed. Ano 2012, São Paulo. Macmillan do Brasil. 136 páginas. ISBN é 978-85-7418-859-1. 2012.

BIBER, Douglas; REPPEN, Randi (Ed.). **The Cambridge handbook of English corpus linguistics**. Cambridge University Press, 2015.

BOUHNİK, Dan; DESHEN, Mor. (2014). WhatsApp goes to school: Mobile instant messaging between teachers and students. *Journal of Information Technology Education: Research*, 13, 217-231. Retrieved from <http://www.jite.org/documents/Vol13/JITEv13ResearchP217-231Bouhnik0601.pdf>

BOULTON, Alex. Looking (for) empirical evidence of data-driven learning at lower levels. Barbara Lewandowska-Tomaszczyk. *Corpus Linguistics, Computer Tools, and Applications: State of the Art.*, Frankfurt: Peter Lang. *Lodz Studies in Language.*, pp.XX, 2008.

BOULTON, Alex. Learning outcomes from corpus consultation. M. Moreno Jaén, F. Serrano Valverde & M. Calzada Pérez. *Exploring New Paths in Language Pedagogy: Lexis and Corpus-Based Language Teaching.*, Equinox, pp. 129-144, 2010.

BROWN, H. Douglas. *Principles of language learning and teaching.* 1994.

CHAMBERS, Angela. *Integrating corpus consultation in language studies.* 2005.

CHAN, T.-W., ROSCHELLE, J., HSI, S., KINSHUK, SHARPLES, M., BROWN, T., PATTON, C., CHERNIAVSKY, J., PEA, R., NORRIS, C., SOLOWAY, E., BALACHEFF, N., SCARDAMALIA, M., DILLENBOURG, P., LOOI, C.-K., MILRAD, M., & HOPPE, U. (2006). One-to-one technology-enhanced learning: An opportunity for global research collaboration. *Research and Practice in Technology Enhanced Learning*, 1 (1), 3–29.

CHENG, Winnie. What can a corpus tell us about language teaching. **The Routledge handbook of corpus linguistics**, p. 319-332, 2010.

CONRAD, Susan. Will corpus linguistics revolutionize grammar teaching in the 21st century?. **Tesol Quarterly**, v. 34, n. 3, p. 548-560, 2000.

CONZETT, Jane. **Teaching collocation: Further developments in the lexical approach.** Hove: Language Teaching Publications, 2000.

Corpus of Global Web-Based English (GloWbE). Available at <<https://corpus.byu.edu/glowbe/>>. Accessed in Apr. 1st 2018.

Corpus of Contemporary American English. Available at <<https://corpus.byu.edu/coca/>> Accessed in Apr. 1st 2018.

COUNCIL OF EUROPE. COUNCIL FOR CULTURAL CO-OPERATION. EDUCATION COMMITTEE. MODERN LANGUAGES DIVISION. **Common European Framework of Reference for Languages: learning, teaching, assessment**. Cambridge university press, 2001. Available at <https://www.coe.int/t/dg4/linguistic/source/framework_en.pdf>. Accessed in May 8th 2018

DE ALMEIDA MATTOS, Andrea Machado. Novos letramentos, ensino de língua estrangeira e o papel da escola pública no século XXI. **Revista X**, v. 1, n. 1, 2011.

DE JONG, F. AND VAN HOUT-WOLTERS, B. (1994) Knowledge construction, self-regulation and processoriented instruction. In: de Jong, F. and van Hout-Wolters, B. (eds.), op. cit., 3–12.

DE OLIVEIRA, Janaina Henriques. O uso de advérbios intensificadores na escrita de aprendizes brasileiros de inglês como L2: um estudo baseado em corpus. 2015.

DUTRA, Deise P. Conscientização linguística com base em corpora online. **Intercâmbio. Revista do Programa de Estudos Pós-Graduados em Linguística Aplicada e Estudos da Linguagem. ISSN 2237-759X**, v. 20, 2009.

DUTRA, Deise P.; SILERO, Rejane P. Descobertas linguísticas para pesquisadores e aprendizes: a Linguística de Corpus e o ensino de gramática. **Revista Brasileira de Linguística Aplicada**, v. 10, n. 4, 2010.

DUTRA, Deise Prina; GOMIDE, Andressa Rodrigues. Compilation of a University Learner Corpus. **BELT-Brazilian English Language Teaching Journal**, p. 21-33. 2015

DUTRA, Deise Prina; GOMIDE, Andressa; OLIVA, Katherine; GUEDES, Annallena. Corpus de aprendizes do Inglês sem Fronteiras: caminhos para compreender a interlíngua de alunos universitários brasileiros. In Sarmento, Simone, Abreu-e-Lima, Denise Martins de, & Moraes

Filhos, Waldenor Barros. Do Inglês sem Fronteiras ao Idiomas sem Fronteiras: na construção de uma política linguística para a internacionalização. p. 151-172. 2016

FRANKENBERG-GARCIA, Ana. Integrating corpora with everyday language teaching. **Input, Process and Product. Development in Teaching and Language Corpora. Brno: Masaryk University Press**, p. 36-53, 2012.

GILMORE, Alex. Using online corpora to develop students' writing skills. **ELT Journal**, v. 63, n. 4, p. 363-372, 2009.

GRANGER, Sylviane; TRIBBLE, Chris. Learner corpus data in the foreign language classroom: Form-focused instruction and data-driven learning. 1998.

Granger, S. 2012a. "How to use foreign and second language learner corpora". In A. Mackey & S. G. Gass (Eds), *A Guide to Research Methods in Second Language Acquisition*. Malden: Basil Blackwell, 7–29.

GRANGER, Sylviane. Contrastive interlanguage analysis: A reappraisal. **International Journal of Learner Corpus Research**, v. 1, n. 1, p. 7-24, 2015.

HYLAND, Ken. **Second language writing**. Ernst Klett Sprachen, 2004.

JOHNS, Tim. Micro-concord: A language learner's research tool. **System**, v. 14, n. 2, p. 151-162, 1986.

JOHNS, Tim. Whence and whither classroom concordancing. **Computer applications in language learning**, p. 9-27, 1988.

JOHNS, Tim. (1991) 'From Printout to Handout: Grammar and Vocabulary Teaching in the Context of Data-driven Learning', **English Language Research Journal** 4: 27–45.

JOHNS, Tim. 19 From printout to handout: Grammar and vocabulary teaching in the context of Data-driven Learning. **Perspectives on pedagogical grammar**, v. 293, 1994.

KHANSIR, Ali Akbar. Error analysis and second language acquisition. **Theory and practice in language studies**, v. 2, n. 5, p. 1027, 2012.

LEHEKA, Tomas. Handbook of Pragmatics 2015. DOI: 10.1075/hop.19.col2 © 2015 John Benjamins Publishing Company

LEMKE, J. (1998) Metamedia literacy: transforming meanings and media. In: Reinking, D., McKenna, M., Labbo, L. and Kieffer, R. (eds.), Handbook of Literacy and Technology: Transformations in a Post-Typographic World. Mahwah, NJ: Lawrence Erlbaum, 283–301.

LEWIS, Morgan. There is nothing as practical as a good theory. **Readings in Methodology**, p. 167, 2000.

Linguee. Available at <<https://www.linguee.com.br/>>. Accessed in Apr. 1st 2018.

MAZON, Marcelo. BLENDED LEARNING E OS DESAFIOS NO ENSINO SUPERIOR Proposta da sala de aula invertida mediada pelas TIC's. **Seminário de Pesquisa, Pós-Graduação e Inovação**, 2016.

MCCARTHY, Michael; MCCARTEN, Jeanne; SANDIFORD, Helen. **Touchstone**. Cambridge, UK: Cambridge University Press, 2005.

MWAKAPINA, Job W.; MHANDENI, Abdulkarim S.; NYINONDI, Onesmo S. Whatsapp mobile tool in second language learning: opportunities, potentials and challenges in higher education settings in Tanzania. 2016.

NASCIMENTO, Roseli Gonçalves do; BEZERRA, Fábio Alexandre Silva; HEBERLE, Viviane Maria. Multiletramentos: iniciação à análise de imagens. **Revista Linguagem & Ensino**, v. 14, n. 2, p. 529-552, 2012.

New Learning Online, 2017. Available at: <<http://newlearningonline.com/new-learning/chapter-1-new-learning/kalantzis-and-cope-a-charter-for-change-in-education/>>. Accessed in Feb. 26th 2017.

O'KEEFFE, Anne; MCCARTHY, Michael; CARTER, Ronald. **From corpus to classroom: Language use and language teaching**. Cambridge University Press, 2007.

O'SULLIVAN, Íde. Enhancing a process-oriented approach to literacy and language learning: The role of corpus consultation literacy. **ReCALL**, v. 19, n. 3, p. 269-286, 2007.

OTERO, N., MILRAD, M., ROGERS, Y., SANTOS, A.J., VERÍSSIMO, M. AND TORRES, N. (2011) 'Challenges in designing seamless-learning scenarios: affective and emotional effects on external representations', *Int. J. Mobile Learning and Organisation*, Vol. 5, No. 1, pp.15–27.

REPPEN, Randi. **Using corpora in the language classroom**. Cambridge University Press, 2010.

SALBEGO, Nayara; HEBERLE, Viviane Maria; DA SILVA BALEN, Maria Gabriela Soares. A visual analysis of English textbooks: Multimodal scaffolded learning. **Calidoscópio**, v. 13, n. 1, p. 5, 2015.

SCHMITT, N. Celce-Murcia. 2002. An overview of applied linguistics. **An Introduction to Applied Linguistics**, p. 1-16.

SINCLAIR, J. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

SINCLAIR, John McHardy (Ed.). **How to use corpora in language teaching**. John Benjamins Publishing, 2004.

SONBUL, Suhad; SCHMITT, Norbert. Direct teaching of vocabulary after reading: is it worth the effort?. **ELT journal**, v. 64, n. 3, 2009, p. 253-260.

SWALES, John M.; FEAK, Christine B. **Navigating academia: Writing supporting genres**. University of Michigan Press, 2011.

TARTONI, Marlei Rose Renzetti. A linguística de corpus no ensino de inglês: um experimento com atividades com "to" e "for". 2012.

VYGOTSKY, L.S. (1978) *Mind in Society*. Cambridge, MA: Harvard University Press

WILLIS, Dave. *The lexical syllabus*. 1990.

WONG, Lung-Hsiang; MILRAD, Marcelo; SPECHT, Marcus (Ed.). **Seamless learning in the age of mobile connectivity**. Singapore: Springer, 2015.

XIE, Fang; JIANG, Xue-mei. Error analysis and the EFL classroom teaching. **Online Submission**, v. 4, n. 9, p. 10-14, 2007.

XIAO, Richard Zhonghua. What can SLA learn from contrastive corpus linguistics? The case of passive constructions in Chinese learner English. **Indonesian JELT**, v. 3, n. 1, p. 1-19, 2007.

APPENDIX 1 – First questionnaire

SoP Draft

* Required

1. Nome Completo *

2. Idade *

• Captura Retangular

Mark only one oval.

- Menos de 18 anos
- De 18-25 anos
- 26-35 anos
- 36-45 anos
- 46-55 anos
- 56-65 anos
- 66 anos ou mais

3. Gênero *

Mark only one oval.

- Feminino
- Masculino
- Outros
- Prefiro não declarar

5. Número de Matrícula: *

• Captura Retangular

6. Curso de Graduação ou Pós-graduação: *

7. Grau máximo de escolaridade: *

Mark only one oval.

- Cursando Graduação
- Graduação Completa
- Cursando Pós-graduação stricto sensu (Mestrado)
- Pós-graduação stricto sensu (Mestrado) - completo
- Cursando Pós-graduação stricto sensu (Doutorado)
- Pós-graduação stricto sensu (Doutorado) - completo
- Outro

8. Há quanto tempo você estuda inglês? **Mark only one oval.*

- Nunca estudei
- Menos de 1 ano
- 1 ano ou mais, mas menos de 2 anos
- 2 anos ou mais, mas menos de 5 anos
- 5 anos ou mais, mas menos de 10 anos
- 10 anos ou mais

9. Você já esteve em algum país de língua inglesa? **Mark only one oval.*

- Não
- Sim. Por mais de 1 mês e menos de 6 meses
- Sim. Por mais de 6 meses e menos de 1 ano
- Sim. Por mais de 1 ano

10. Qual sua língua materna? *

11. E-mail: *

Termo de Consentimento Livre e Esclarecido

O Sr. (a) está sendo convidado (a) como voluntário (a) a participar da pesquisa "Dictionaries and Online Corpora as writing tools", integrante do projeto "Traços linguísticos do discurso acadêmico: um estudo de corpora de aprendiz e de textos científicos especializados" e coordenado pela Profa. Deise Prina Dutra. Pedimos a sua autorização para a coleta e análise de seus textos escritos em inglês. A utilização de seus textos está vinculada somente a este projeto de pesquisa ou se Sr. (a) concordar em outros futuros. Nesta pesquisa pretendemos descrever o uso da língua inglesa por aprendizes brasileiros universitários de inglês bem como comparar esse uso com traços linguísticos presentes em artigos acadêmico-científicos. Para esta pesquisa os participantes farão tarefas na disciplinas Inglês para Fins Acadêmicos, utilizando seus conhecimentos prévios e sendo incentivados a usar ferramentas que podem auxiliar a escrita. Os desconfortos dos participantes são mínimos, podendo se sentir pressionados por saberem que seus textos farão parte de uma pesquisa. Esses desconfortos serão minimizados, pois as atividades de coleta de dados são atividades comumente feitas em sala de aula. Além disso, garanto que identificação de nenhum dos participantes será divulgada. A pesquisa contribuirá para o aumento do conhecimento a respeito dos processos de aquisição favorecidos com a utilização de corpora eletrônicos, podendo beneficiar outros aprendizes de inglês. Para participar deste estudo o Sr. (a) não terá nenhum custo, nem receberá qualquer vantagem financeira. O Sr. (a) terá o esclarecimento sobre o estudo em qualquer aspecto que desejar e estará livre para participar ou recusar-se a participar e a qualquer tempo e sem quaisquer prejuízos, pode retirar o consentimento de participação na pesquisa, valendo a desistência a partir da data de formalização desta. A sua participação é voluntária, e a recusa em participar não acarretará qualquer penalidade ou modificação na forma em que o Sr. (a) é atendido (a) pela pesquisadora, que tratará a sua identidade com padrões profissionais de sigilo. Os resultados obtidos pela pesquisa estarão à sua disposição quando finalizada. Seu nome ou o material que indique sua participação não será liberado sem a sua permissão. O (A) Sr. (a) não será identificado (a) em nenhuma publicação que possa resultar.

Os dados e materiais utilizados na pesquisa ficarão arquivados com o pesquisador responsável por um período de 5 (cinco) anos na sala 4111 da Faculdade de Letras da UFMG e após esse tempo serão destruídos.

Os pesquisadores tratarão a sua identidade com padrões profissionais de sigilo, utilizando as

informações somente para fins acadêmicos e científicos.

Fui informado (a) dos objetivos, métodos, riscos e benefícios da pesquisa de maneira clara e detalhada e esclareci minhas dúvidas. Sei que a qualquer momento poderei solicitar novas informações e modificar minha decisão de participar se assim o desejar.

12. Por favor clique em uma das opções: *

Mark only one oval.

- Concordo que os meus textos escritos em inglês sejam utilizados somente para esta pesquisa.
- Concordo que os meus textos escritos em inglês possam ser utilizados em outras pesquisas, mas serei comunicado pelo pesquisador novamente e assinarei outro termo de consentimento livre e esclarecido que explique para que será utilizado o material.
- Não concordo em participar da pesquisa.

Nome completo do Pesquisador Responsável: Deise Prina Dutra

Endereço:

Rua Expedicionário Nilo Morais Pinheiro, 660, casa 16

CEP 31310-060, Belo Horizonte - MG

Telefones: (31) 3568-0836; (31) 99106-8331

E-mail: deisepdutra@gmail.com

CPF: 091599848-30

RG: MG 16508030

Mestranda responsável: Katherine Oliva

Em caso de dúvidas, com respeito aos aspectos éticos desta pesquisa, você poderá consultar:

COEP-UFMG - Comissão de Ética em Pesquisa da UFMG

Av. Antônio Carlos, 6627. Unidade Administrativa II - 2º andar - Sala 2005.

Campus Pampulha. Belo Horizonte, MG – Brasil. CEP: 31270-901.

E-mail: coep@prpq.ufmg.br. Tel: 34094592.

13. Escreva aqui o seu Draft de Statement of Purpose *

Research questionnaire IFA 1:

14. Which tools did you use while you were writing your SoP Draft? *

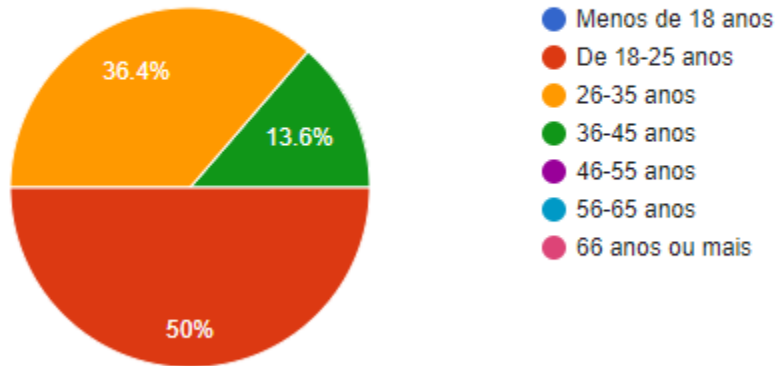
Check all that apply.

- Google Tradutor
- Online monolingual dictionary (English)
- Paper version of monolingual dictionary (English)
- Online bilingual dictionary (English – Portuguese)
- Paper version of bilingual dictionary (English – Portuguese)
- COMPARA
- Linguee
- Coca Corpus
- Glowbe corpus
- Other: _____

APPENDIX 2 - Responses

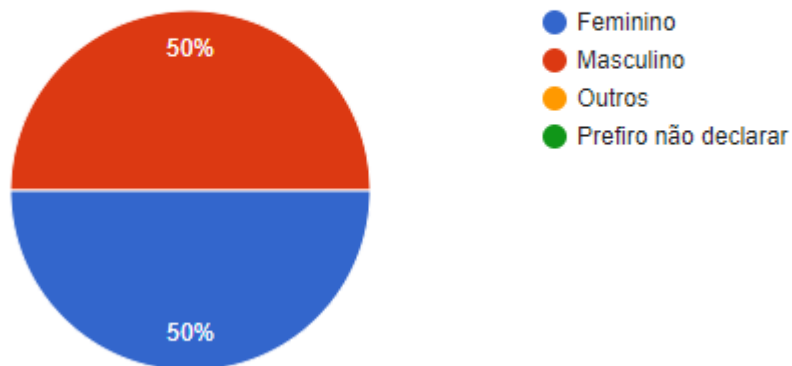
Idade

22 responses



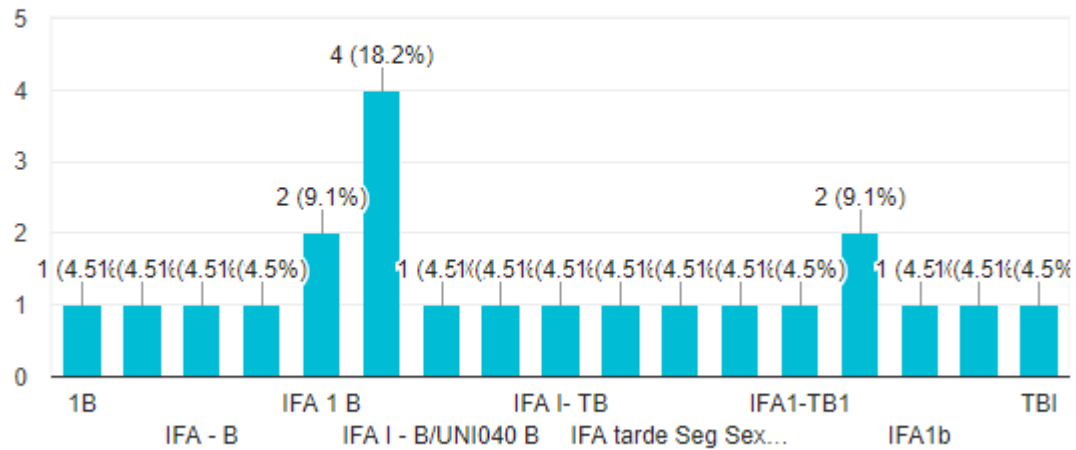
Gênero

22 responses



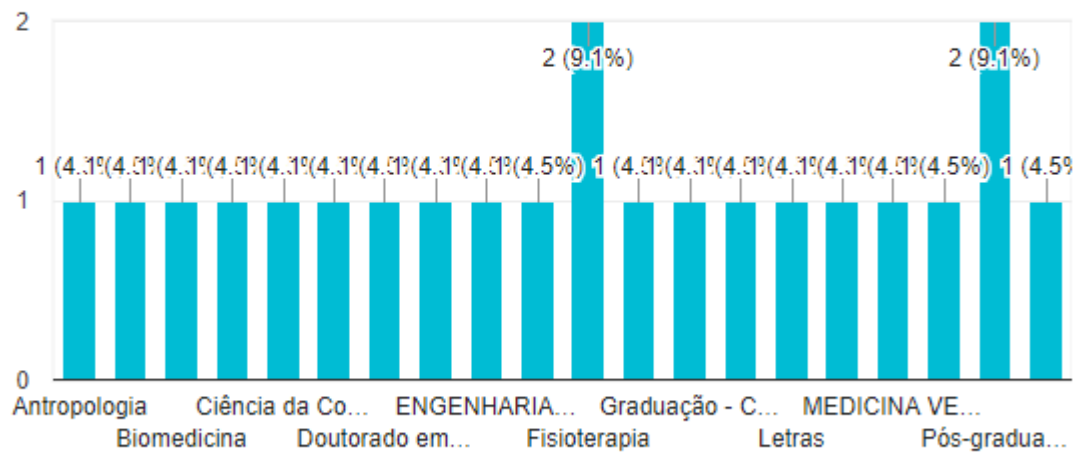
Turma IFA:

22 responses



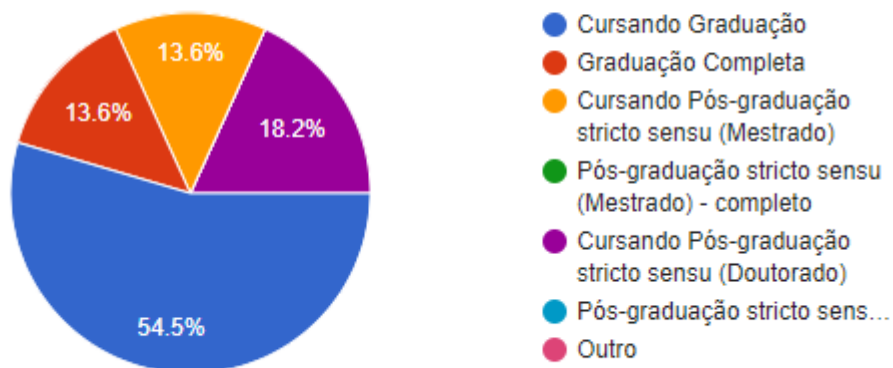
Curso de Graduação ou Pós-graduação:

22 responses



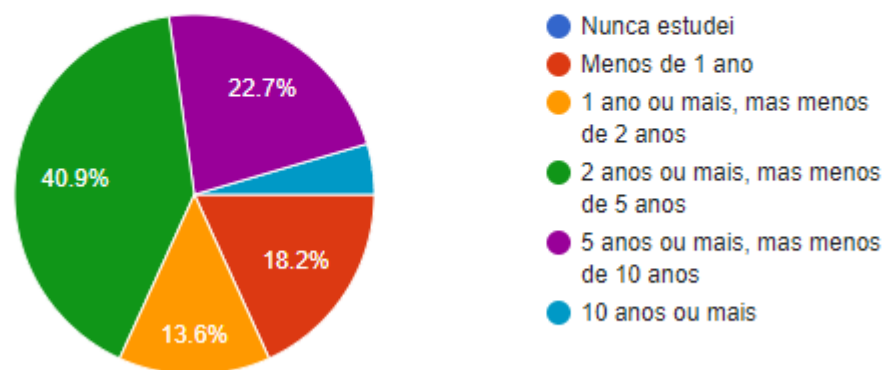
Grau máximo de escolaridade:

22 responses



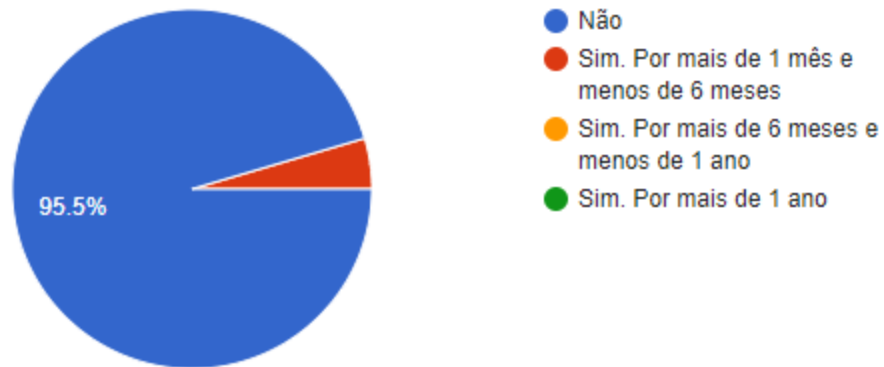
Há quanto tempo você estuda inglês?

22 responses



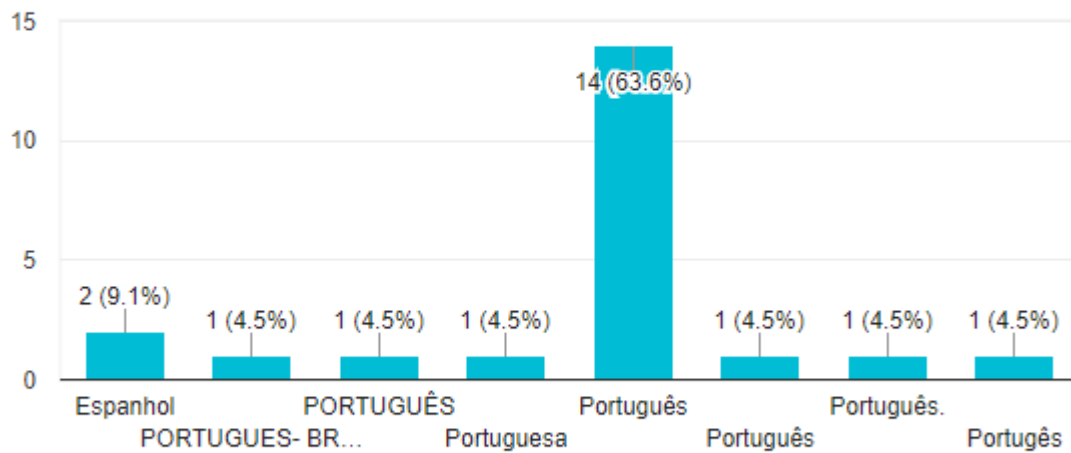
Você já esteve em algum país de língua inglesa?

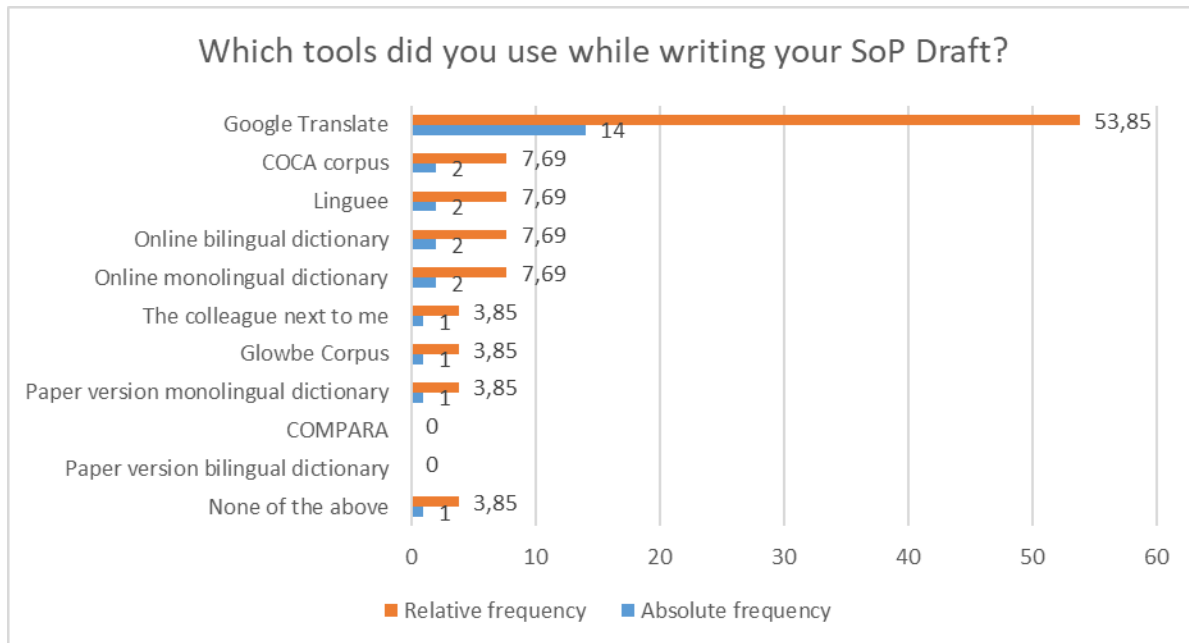
22 responses



Qual sua língua materna?

22 responses





APPENDIX 3 – SoP Drafts

TEXT 1

My name is xxxx **and a study english for a any years**. I have so much difficulty with progres in **english**, be in speaking or comprehension. I would **lik** to **know try make a exchange** with on university **on USA** or other english speaker country **for improve** my knowledge and my conversation. To me be a part of exchange create a new world and new vision of possibilities for my job.

Im a researcher in a dance area. My research is about de **middle** of learning **between observation** and reproduce, indeed, my research is about 鍍he copy • This area relation with a neural process happened when someone look or watching a gestural movement and reproduce in you own body.

My main purpos is to share my research and found other connections in this area with **others researchers** of dance an improve my knowledge in my research and my language.

TEXT 2

Hi, my name is XXX Mascarenhas Diniz. I'm graduated in Building Technician. I'm student of Civil Engineering **on UFGM**, I'm on 3th term. I have the dream of studying in other countries to get to know other cultures and see other ways of solving the problems of construction. But I have difficulty with English, and I intend to improve it by **doing an exchange**. **I would like know Australia** or Japan. In this way, I will have more opportunities for my job.

TEXT 3

I am single, **i borne** in Porteirinha, inside the country, now **i** live in Belo Horizonte Capital of Minas gerais in Brazil, live in here for studying Biology at University of Minas Gerais, **i** am in period seven of nine. **I am loved Biology**, more the part of human health, the mechanism of funcionament of the brain, my interest in Biology is in Neuroscience, the part of physiology, **how happens interactions** between the neurons. My hobbies are, read books, all types, since classics until current bestsellers, and mainly lecture about science. Before **i** graduate, **i** apply for the abroad Master's Degree, **i dream go** to Canada or Germany, and continue to **studs** until PhD Biology or similar areas .

During my graduation, in integrate a team of scientific initiation in Laboratory of Angiogenesis and Stem-Cell, in this laboratory, **i am research about** influence of MAS receptor, belonging to the Angiotensin complex in ischemic murine models. I am participating in more than four projects in the Laboratory, helping my team in other **research lines**. **I work since twelve years**, because in city **when i borne**, they accept all the kids, my first job how was it delivered the bread, after **i** work in restaurant, in supermarket, in ice cream shop, my last job was **em** 2013, in planning strategies for a company the call center.

TEXT 4

Dear professor, My name is XXX. I am a PHD student of arts, I am searching the negative charge of the shadow in the myths, arts, and fine arts. I chose this field of study because I want to know why the shadow is the principal esthetic element of the fine arts mise en scène **to can** arouse fear and panic. Also, I want to study the relation between the shadow and the sound in the expressionist films. For this reason, I want to develop this part of my search in your prestigious institution.

I am doing this search **whit** the help of my tutor PHD XXXX for four years. **We public papers** of this subject in Revista Trama (Rio The Janeiro) and Avanca book of papers (Portugal), also I **participate in** GRUDES conference in Curitiba, also **I will public** a paper in the GRUDES book of papers.

TEXT 5

I am a Bazilian, public employee, graduated in **Information Science** and Fashion Design student. **I am always have interest in draw**, make handicraft like accessories and observing how the fashion influences the **people life**.

At the moment **I am doing the undergraduate** in fashion design, and I am interested in modeling, printworks, accessories and the influence of fashion design in the daily life of the society.

I would like to unite and **apply Information Science and Fashion Design in** my professional life.

TEXT 6

In the college I figured out why I like science and **what makes me to keep continue** my studies after I graduated. During the time I was **in the college**, I applied to **participate in** a laboratory of sistematic of plants, I had an excelent experience working with plants and I realized that I am a botanist. To follow my passion in this work, I decided that I would like to improve my knowledge.

TEXT 7

Statement of Purpose **I would to study** in this university because my interests in Biomedical Science started when I was reading about DNA and Human Genoma Project. **During my graduate** in Information Systems, I studied data bases xxx and now I am studying xxx

TEXT 8

When I decided to course Letras, during my classes I started **to being** very excited about the studies in language acquisition. **This interesting intensify** because (...) During the first semester of 2017, I started to **work in a project** that was dedicated to **knew professor**, to be their first jobs and know about their area a little more. In this project we had the opportunity to study and see a little how was the process of second language acquisition. (...) Now I'm taking classes of Phonology, that will help me a lot in my pronunciation and with..., last semester I **toked** class of Oral Expression that was **a discipline that to help** us with our English skills, we had to learn about how **was the processes** of **listening a different language**, **based in** ... , to **help us to discuss** about this topic in the classes. This help me to **to deciced** CHICO university because you have a group **related of** education.

TEXT 9

I started my undergraduate studies at the Faculty of Economics from UFMG, **in Economic field** related to International relations. My main interest is to **reconcile in the best way possible** my knowledge about Public Law and Economics field.

My professional aspirations are teachers and professors that I had during my High School and undergraduate time, who I had the pleasure to live with and **divide academic experiences**. I was fortunate to have **good intern** and excellent professors during my academic life.

After two year's work experience as an intern on Public Ministry of Justice (on "Fazenda Pública") at my first graduation, I gain some practical experience working with Public Law and I could see how important **the economic** was to Public Administration, especially macroeconomics and microeconomics, now I **have interest to know** more about public politics and how to make **a word a better place** to live.

My love of Economic started during my Post-Graduation Lato Sensu Course on Public Law field, there I met teachers who **introduce to me** a lot **Economics authors**, most of them with divergent perspectives from **economic**. From disciplines like **Fiscal Responsibility** and Financial Law I could understand how was important the economic planning to Public Administration at the same way that was important the correct application of the law. One professor also encouraged us to read Capital in the twenty-first century, a book from Piketty a very famous Economist, I remember clearly the impact reading these book had on me. This interdisciplinary aspect of the subject **helped me to make** my choices.

The main characteristics of both professional areas are to be proactive, analytic, critical not only on professional time, but on free time **either**. I'm a person who likes to discover new things every time, I like to do research by myself about economics and law fields. I also like to read and to **listen people** who have different opinions from me about many varieties of things. **I have interest** to know more about public politics and how to **make a word** a better place to live and at undergraduate level, my interests became more focused.

In selecting the University of Chicago I have been influenced by the way that this University treats their members, how they conduct the career planning and how profound influence of economic researches on a wide diversity of areas.

TEXT 10

My interest for **electrical engineering** began when I started a professional formation course in electrical industry at SENAI/FIEMG school, in 2008. After that, **in first semester** of 2009, introduced to electrical engineering in high school I realized that it **interestingly qualified** as a subject of both Power Systems and Renewable Energy.

TEXT 11

On my career plan, **i** hope to **work on** Material Science as research science or make my own business. When I accumulate enough experience I will **try follow** what **i** want. Absorbing new technologies is always a pleasure to me ,so I chose to study further. Studying abroad has been my dream since I was in junior high school. First of all, I wish to get **a undergraduate** on Materials Science and Engineering. While I study in the field of Materials, I hope to

implement my ideas on industry in general, **including find** methods **more clean** to produce. I chose the MIT because he is the best university in the world in this field.

TEXT 12

I am a Phd student in the Biochemistry Department from University of Minas Gerais, Brazil. I concluded my Master last year and graduated in Biology in 2013 at the same university. **Actually,** I work with Molecular Biology in Immunochemistry of Proteins and my research involves antivenoms. Dr. Carlos Chavez Olórtegui is specialist in antivenoms in Brazil **e** Peru, and I participate in **your** research group. Our lab collaborates with Braga Lab, the lab of Dr. Vânia Braga, Imperial College member. We intend to strengthen our partnership with new experiments that I will **realize** in her lab.

TEXT 13

Since I **began in the Undergraduate** Course, I was sure that my work field will be Sport Science. I started the Physical Therapy course at UFMG in 2015 (rated as the best Physical Therapy course in Brazil) and always **focusing in** that area, searching and reading as much as I can, and getting in tutor programs for undergraduate students. I started my professional life even before coming to the University. In 2011, **I went to my first job**, as a Graphic Designer, and worked until 2014 when I came to UFPA to get in a Physical Education School, in which I get out in 2015 and started in Physical Therapy in UFMG.

APPENDIX 4 - Workshop

Slide 1

Integrating Corpora with Language Classroom

Katherine Oliva
PosLin
Workshop at IFA - I
September 2017

Slide 2

What is “corpus”?

- It is a collection of texts. Corpora is the Latin plural of corpus. Slide 3
- **COCA** - Corpus of Contemporary American English
 - It shows us common language structures through real world context
 - Textbooks may not provide everything you need on your own necessities
 - Search for words that frequently occur together
 - It contains over 520 million words and is frequently updated
 - It has many genres divided among:
 - Spoken (from radio and TV)
 - Fiction
 - Magazine
 - Newspaper
 - Academic

Let's figure this out!!!

Slide 4

1. Access <http://corpus.byu.edu/coca/> and **register**.
2. In the **DISPLAY** section click on **List**
3. Type blame in the SEARCH STRING box and click "see frequency by section".
4. Separate blame in two categories: a) **Noun** b) **verb**

What was the criteria you used to make that classification?

Slide 5

KWIC - Parts of speech

It shows how a word can be used, how it appears in a sentence.

Each part of speech has a different color.

Let's check it out?!

Corpus of Contemporary American English

SEARCH FREQUENCY CONTEXT ACCOUNT

Slide 6

CLICK FOR MORE CONTEXT	SAVE LIST	CHOOSE LIST	CREATE NEW LIST	SHOW DUPLICATES
1 1999 ACAD PSAJournal	A B C	want memories and photographs of the things ve described ;	book trip to Botswana . Plan on a long flight over ,	
2 1991 SPOK CNN_King	A B C	, what you can do- KING : That ll be a	book week from Tuesday . Mr. LESKO : laughs That 's right	
3 2002 ACAD SocialStudies	A B C	many valuable resources to those who harvest them ; Cherry 's	book about is impose in its approach . Cherry 's rainforest	
4 2005 NEWS Chicago	A B C	attention to such behavior because I was planning to write a	book about religion movements started by manic prophets . But I	
5 2012 SPOK NPR_FreshAir	A B C	Grabell is an investigative reporter for ProPublica ; His new	book about the economic stimulus plan is called " Money Well Spent ?	
6 2006 SPOK Fox_Sunday	A B C	possibility of running for president in 2008 ; in a new	book about you your wife , Lilibet , is quoted as being	
7 2008 NEWS AssocPress	A B C	// " People would often ask ; ! is this	book about you " I did n't want to talk about that	
8 2006 FIC SouthernRev	A B C	brim until that creeps me out such that I close my	book legally and set up . thinking this is turning into a picaresque	
9 1992 NEWS Chicago	A B C	# The unusual campaign got a lot of attention for the	book and contributed its success - 27,000 copies sold (or given	
10 1996 NEWS Atlanta	A B C	troubling medical report , Dan Sweat took his eyes off his	book and stared in silence , at his wife , # "	
11 2012 SPOK CNN	A B C	's the plan . So he put it together in a	book and that is a smart idea for those who are undecided .	
12 1996 FIC Bk:DeepEndOcean	A B C	he went back , Reese looked Kilgore up in the phone	book and there he was , Thomas K. Kilgore , M.D. So that	
13 2009 ACAD Writer	A B C	her he was an author , gave the name of his	book and Web site ; and commented on a few of the magazine	
14 2006 NEWS Atlanta	A B C	picking on his son -- a kindergartner with a new Spider-Man	book bag is the end of the school year , Noah might	
15 2006 NEWS Chicago	A B C	that at the time . " # Perhaps Oprah adored the	book because is had striking similarities to The Color Purple , a book	
16 2002 MAG Essence	A B C	THE BEAUTY MECCAS THAT SHOULD BE LISTED IN YOUR STYLE BUCKET	BOOK Best Mail-Order Catalog # Blissout (888) 243-8825 : This	
17 2004 ACAD InfoSystems	A B C	's standard package . Instructors may also want to acquire the	book CAATTs and Other Beasts for Auditors (Coderre 2001a) . For	
18 1998 NEWS WashPost	A B C	at Harvard Law School and the author of a widely read	book called is getting to Yes , " the major point of which	
19 1994 SPOK CBS_EyeToEye	A B C	of Western Ontario in Canada . He is just written a	book called is Race , Evolution and Behavior , " and his theories	
20 1999 SPOK CBS_SatMorn	A B C	years ago . Their story is the subject of a new	book called is The Endurance : Shackleton 's Legendary Antarctic	
21 1998 NEWS Denver	A B C	" In the order in which it was published , The	book came out in " islands of story , " Lamb said .	
22 2012 ACAD Ref&UserServQ	A B C	peer-reviewed journal article , conference proceedings ; and	book chapters range from 1928 to the present . # LEGISLATION AND	
23 2008 MAG USAToday	A B C	a two dollar trade deficit with Taiwan . Yes , the	book comes back here and retails for \$24.95 . The value added is	
24 1997 ACAD NaturalHist	A B C	into pieces to smuggle it past the soldiers ; Although the	book contains photos of the author and his son climbing , camping ,	
25 1993 SPOK PBS_Newshour	A B C	There 's no money involved . MR-KURTZ : They want a	book contract is MS-CHAVEZ ; They may want a book contract .	
26 2007 ACAD Mercury	A B C	" you " will do this and that ; # The	book depicts an explicit character rednecks boy as his well-dressed	
27 2006 NEWS Chicago	A B C	UP # The N.Y. Post , plugging Boze Hadleigh 's	book Disc is Tell ; Stars Talk About Each Other , quotes Jennifer	
28 1994 NEWS CSMonitor	A B C	but a very few portraits on the medals illustrated in the	book do and on the other (the reverse) a scene	

Slide 8

CHART

It shows graphs representing frequency of a term in a variety of texts.

BUT x HOWEVER

(HOWEVER is more frequent in **academic**)

You can also see **areas of study** by clicking on “academic”, “fiction” etc...

Slide 9

Collocates

1. It helps us understand the meaning and usage of the word we are investigating.

2. MAKE x DO

Make homework X

do homeowrok

????????????????

NOW YOU!

1. Type the expression look forward to in the 'WORD(S)' box
2. Click on COLLOCATE and select the spam 0 to the left and 1 to the right
3. In the POS LIST's box, select the option **verb.ALL**

Pair work!!!

Slide 10

1. Complete the expressions below with the given prepositions.

- a. Skilled __
- b. concentrate __
- c. rely ____
- D. fascinated ____

2. Do the same with the expressions below and analyse the results. Select the spam according to your need.

Slide 11

- a. enjoy + verb
- b. stop + verb (spam 1 and spam 2 - to the right)
- c. preposition + Monday
- f. pleased + verb (spam - 2 words to the right)
- h. Interested + preposition
- i. Help + preposition

Slide 12

Have you heard of GloWbE?

The Corpus of Global Web-Based English (GloWbE)

1.9 billion words from 1.8 million web pages in 20 different English-speaking countries.

Created by Mark Davies of Brigham Young University, released in 2013.

GloWbE (pronounced like "globe") is related to Corpus of Contemporary American English (COCA) and Corpus of Historical American English (COHA). Together, these three corpora allow researchers to examine variation in English -- by dialect, genre, and over time -- in ways that are not possible with any other large corpora of English.

Slide 13

1. Go on <http://corpus.byu.edu/glowbe/> .
2. In the **DISPLAY** section click on **CHART**
3. Type color in the SEARCH STRING box and click "see frequency by section". In which country does the word present the highest frequency?
4. Type colour in the SEARCH STRING box and click SEARCH. In which country does the word present the highest frequency?
5. Type *colo*r* in the SEARCH STRING box and click SEARCH. Compare the result with your analysis of the two previous tasks.
6. Do the same with gray, grey.

Slide 14

Activity adapted from Bixby & Scanlon (2013)

Do you know the possible prefixes of the words below???

a. appropriate	b. certain	c. conscious	d. convenient
e. dependent	f. experienced	g. fortunate	h. intelligent
i. legal	j. mature	k. necessary	l. perfect
m. practical	n. precise	o. sufficient	p. resistible

Slide 15

Activity adapted from Bixby & Scanlon (2013)

Choose List and write the words preceded by*

- One of the techniques used to write paraphrase is to add **prefixes to change the meaning** of a word to the opposite. Using the wildcard *, access <http://corpus.byu.edu/coca/> and identify the correct prefixes for the words below.

a. appropriate	b. certain	c. conscious	d. convenient
e. dependent	f. experienced	g. fortunate	h. intelligent
i. legal	j. mature	k. necessary	l. perfect
m. practical	n. precise	o. sufficient	p. resistible

Slide 16

When you write, are you repetitive?

Do you want to write the same word again and again?

How about SYNONYMS ????

HAVE YOU THOUGHT OF THAT ???

How can I find synonyms? Can I use COCA for that?

YES!!!!!!

Slide 17

HOW ???

On list, type [=word] on the research box.

But be careful! Depending on what you want to say, not all the words could work!..

So it is very important that you read the context where that specific word appears.

Useful links

<https://corpus.byu.edu/coca/>

<https://www.youtube.com/watch?v=sCLqRTIxG0Y>

<https://www.youtube.com/watch?v=KP-7thiUnLM>

https://www.youtube.com/watch?v=t_SxpfjPo_o

Activities created by Andressa Gomide, Katherine Oliva and Michel de Souza

APPENDIX 5 – Activities

Integrating Corpora with Language Classroom

Mestranda Katherine Oliva
Orientadora Deise Dutra
PosLin
Workshop at IFA - I
September 2017

Interferência linguística

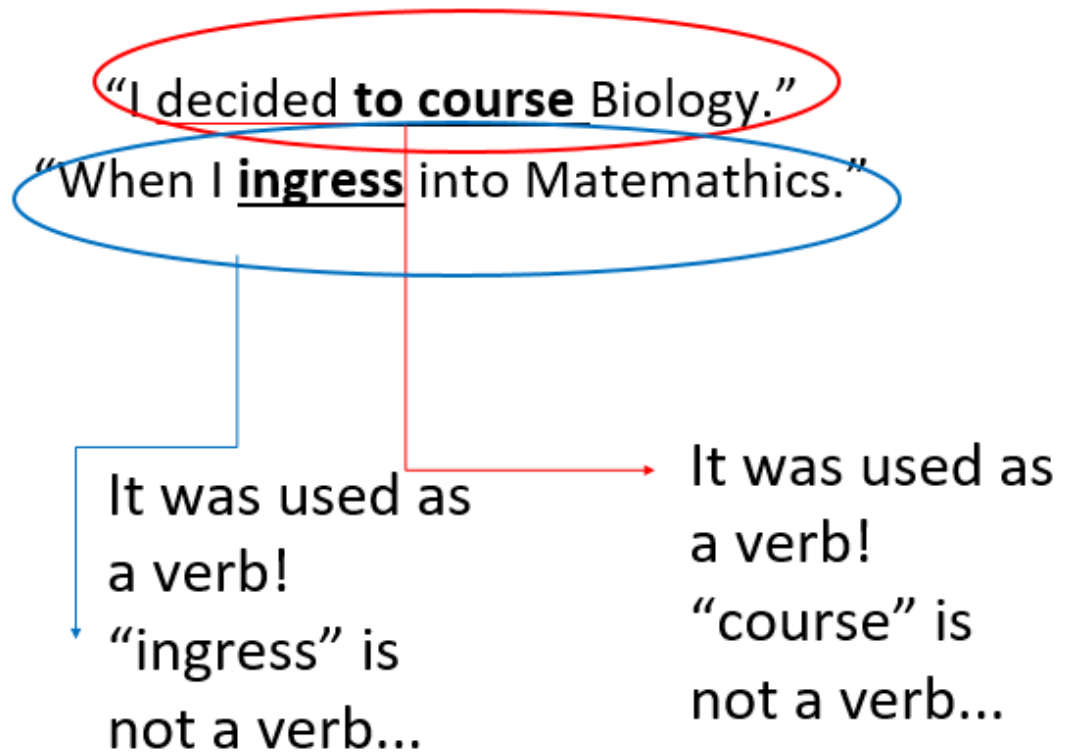
“I decided to course Biology.”
“When I ingress into Matemathics.”

Do these words exist in English???

Let's check!

OH YEAH!!!! THEY DO!!!! But....

Interferência linguística



Spelling and Lexical problems

- Go on COCA
- List
- Type "ENGLISH"
- Now do it with "BRAZILIAN"
- Now with "AMERICAN"
- Now with "PORTUGUESE"

Spelling and Lexical problems

- What patterns were you able to notice?

CAPITAL LETTERS ALWAYS!!!!!!!

English

Brazilian

American

Portuguese

Spelling and Lexical problems

- Biomechanics
- System Engineering
- Information Science

Is it how they say your major in English?

Let's check????

Spelling and Lexical problems

“Divide experiences”

- Check “divide” or “experiences”.

How???

- **Try List, collocates... Let’s think!**

Spelling and Lexical problems

“Divide experiences”

Share experiences.

Grammar and structure

- Are these correct?

like know

Love eat

Need study

Have write

- Is there anything missing?

Grammar and structure

What are these words?

like know

Love eat

Need study

Have write

Verbs

Grammar and structure

like know

Love eat

Need study

Have write

**Let's check how
verbs occur with
other verbs!!!!**

Grammar and structure

1 – Go on collocates

2- Type *try_v*.

3- Choose verb.aLL

4- choose 1 span to the right

choose 2 span to the right

choose 3...

Grammar and structure







1 – Using KWIC, check sentences with the verb “JOIN”.

- What can you observe?

American English			
FREQUENCY	CONTEXT	ACCOUNT	
	join	a free trade area	# Only such governments can be members of the GNA , or participate in the
	join	a gang at all	make is about which gang to join , or whether to
	join	a group of	, because she was the one who actually invited me to
	join	a group practice here	" Paul Seamars told me . " If I wanted to
	join	a hose connector to	, to one end of a length of PVC pipe .
	join	a much smaller venture	do , " Karanja says . " I 'm going to
	join	a regional group	several talent showcases . Afterwards , he had been advised to
	join	a religion	going to be effectiveness . People practice - ultimately people
	join	a self-help group	If you think you need help , seek grief counseling or
	join	a smaller law practice	environment of the nation 's largest law firms and
	join	a study circuit	study circuit comment on each other 's images every month .
	join	a team of investigators	At the Chimney Rock Archaeological Area , you 'll need to
	join	a tour	just walk around the house and grounds , you have to
	join	a union so they	That 's what 's supposed -- that 's the reason you
	join	action groups such as	held every two months . # ' Patients are encouraged to
	join	Al Shabab , # Hammami	on terrorism violations , including for leaving the US to


















Grammar and structure

2 – Using **Collocates**, type any noun, plural form (ex.: subjects, cultures, people, countries, houses.....) and look for “adjective”. Choose 1 span to the left.

emporary American English      

FREQUENCY CONTEXT HELP

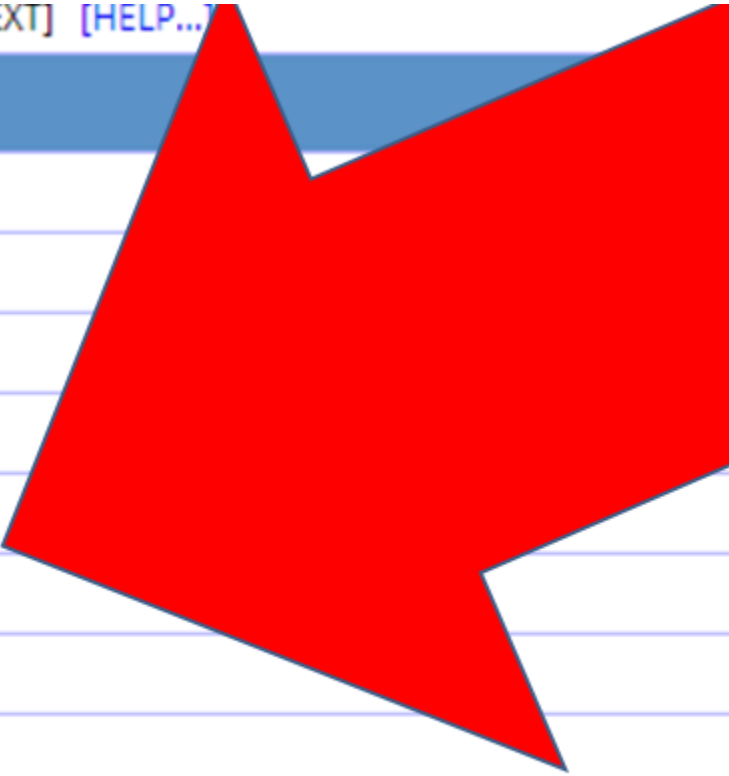
OR SELECT WORDS + [CONTEXT] [\[HELP...\]](#) COMPARE

CONTEXT	FREQ	
OTHER	689	
HUMAN	491	
ACADEMIC	210	
FEMALE	149	
DIFFERENT	137	
MALE	118	
NORMAL	117	
TECHNICAL	115	
FAVORITE	111	
HEALTHY	86	
CERTAIN	80	
EXPERIMENTAL	62	
VARIOUS	61	
LOYAL	59	
ELDERLY	57	
NEW	57	
AMERICAN	54	

SoPs com feedba....docx ^ E

OR SELECT WORDS + [CONTEXT] [HELP...]

CONTEXT
OTHER
HUMAN
ACADEMIC
FEMALE
DIFFERENT
MALE
NORMAL
TECHNICAL
FAVORITE
HEALTHY



Grammar and structure

- What can you observe about the adjectives that follow the noun?

SINGULAR!

Grammar and structure

- How do we use adjectives in English?

ADJECTIVE (always singular) + **Noun**

Grammar and structure

- Which is correct?
 - Subjects different
 - Subject different
 - Different subjects
 - Differents subjects

Grammar and structure

- Which is correct?
 - Subjects different
 - Subject different
 - Different subjects** ←
 - Differents subjects

Articles a/an

“Read a article about engineering.”

Is this correct?

NO

- Go on COCA
- Collocates
- Type “article”
- Choose art.aLL
- 1 span to the left

- Go on COCA
- Collocates
- Type “article”
- Choose art.aLL
- 1 span to the left

“Read an article about engineering.”

APPENDIX 6- Errors selected from the second pilot study

SoP 1	
Problem	Correct
<u>fascinated with things that fly</u>	Fascinated <u>by</u> things that fly
<u>excited to join</u>	

SoP 2	
Problem	Correct
<u>ingress into engineering</u>	
<u>fascinated to understand</u>	
<u>help me to have</u>	

SoP 3	
Problem	Correct
<u>I'm making master student</u>	

<u>many project</u>	
<u>necessity about to improve</u>	
<u>with focus poultry science</u>	
<u>The University promote</u>	
work a lot <u>in research</u>	

SoP 4	
Problem	Correct
<u>study in the Applied Linguistics field</u>	
<u>I participated in a beneficial course</u>	
<u>In the weekends</u>	
<u>focus only in the studies</u>	

APPENDIX 7 – Second questionnaire

Final SoP Version

This SoP will be graded


* Required

What's your name? *

Your answer _____

Write your SOP here: *

Your answer _____

NEXT  Page 1 of 2

Never submit passwords through Google Forms.

Questions about your writing process

The aim of this form is to find out about the tools you used while writing (if you used any)

Quais ferramentas foram usadas por você para melhorar seu texto (segunda versão)? *

- Google Tradutor
- Dicionário monolíngue ONLINE (inglês - inglês)
- Dicionário monolíngue DE PAPEL (inglês - inglês)
- Dicionário bilíngue ONLINE (inglês - português)
- Dicionário bilíngue DE PAPEL (inglês - português)
- COMPARA
- Linguee
- COCA CORPUS
- GLOWBE CORPUS
- Other: _____

No dia-a-dia, ao escrever algo em inglês (e-mail, textos acadêmicos e outros), você utiliza... *

- Google Tradutor
- Dicionário monolíngue ONLINE (inglês - inglês)
- Dicionário monolíngue DE PAPEL (inglês - inglês)
- Dicionário bilíngue ONLINE (inglês - português)
- Dicionário bilíngue DE PAPEL (inglês - português)
- COMPARA
- Linguee
- COCA CORPUS
- GLOWBE CORPUS
- Nenhum
- Other: _____

Quais foram as ferramentas do COCA mais usadas por você para corrigir seu SoP?

- List
- Chart
- Collocates
- KWIC
- Compare
- Nem usei o COCA

A respeito do COCA, você considera... *

- Útil, foi válido ter aprendido a respeito.
- Útil, me ajudou a melhorar meu texto
- Gostaria de aprender mais a respeito
- Vou usar mais vezes
- Não faz diferença na minha vida, pois não uso mesmo...
- Inútil, não me ajuda em nada.
- Other: _____

SUBMIT

Page 1 of 1

APPENDIX 8 – Edited texts

TEXT 1

My name is XXX, I am 38 years old and living in Belo Horizonte, at Minas Gerais State. I study English for 2 years. I am undergraduate in Dance on Escola de Belas Artes da Universidade Federal de Minas Gerais since 2012. Throughout of these time I knew and study a little bit of Psychology Area and I liked so much. Indeed it is a topic very interesting because my Research pass through the neurological process and comportamental psychology.

My Scientific Initiation analysing the process of observation and reproducing, basically, I study the process of “Copy” the movement through observation. Initially I research any specifics Banks of Datas which contain Thesis, Dissertations and Researches about and inside the Art Dance Education with circumscribed cut of the keywords: Copy, Observation, Appropriation and Reproduction. By the other hand I seeking the keywords: Breaking, Hip Hop, Urban Dance and Street Dance, for deeper a little more on researcher. These process help me to improve the knowledge on my field research and pointing for other directions look like a Psychology and Neuroscience, two interesting areas for I sink.

Nowadays I am study hard the English language for to try improve my skills and open new opportunities for Research abroad. I have so much difficulty with progres in English, be in speaking or comprehension. But I hope to try one chance for to do an exchange. To me be a part of exchange create a new world and new vision of possibilities for my study.

TEXT 2

Hi, my name is XXX. I'm graduated in Building Technician. I'm student of Civil Engineering in UFMG, I'm on 3th term. I have the dream of studying in other countries to get to know other cultures and see other ways of solving the problems of construction. But I have difficulty in speaking English, and I intend to improve it by doing an exchange. I would like go to University of Sydney. In this way, I will have more opportunities for my job.

In my academic life, I have participated in a scientific initiation: "Mapeamento dos Residuos solidos gerados pela construção civil em Curvelo-MG"-Mapping of solid waste generated by the civil construction in Curvelo-MG (2015). In addition, I was the director of the academic extension project "Palco a um passo- a arte transformando vidas"-Stage by step - art transforming lives (2014/2015). I also did an internship at the company "Flamer Construções"-Flamer Constructions (2016). With these experiences I have become a much more mature, responsible, extroverted and communicative, agile and innovative person.

Currently I am participating in the selective junior company (EMAS) of my course, in addition I'm involved with a project of scientific initiation "Caminhabilidade e pedestrianismo em Belo Horizonte"-Walkability and Pedestrianism in Belo Horizonte and with a project of academic extension "Metodos de pesquisa e inovação"-Research and innovation methods.

The University of Sydney is famous for its shorter class schedule and greater incentive to research. This kind of grid organization attracts me.

TEXT 3

My name is XXX, I was born in Porteirinha, inside the state, now I live in Belo Horizonte Capital of Minas Gerais in Brazil, live in here for studying Biology at University of Minas Gerais, I am in period seven of nine. I always dreamed of being a biologist, my interest area is Neuroscience. My hobbies are, read books, all types, since classics until current bestsellers, and mainly lecture about science. after graduation I intend to apply for the abroad Master's Degree. My dream is to study in Germany, and continue to study until PhD Biology or similar areas.

During my graduation, in integrate a team of scientific initiation in Laboratory of Angiogenesis and Stem-Cell. In this laboratory I am participating in more than four projects, helping my team in other lines of research. I worked in a Company for communications during 7 years, where I get abilities for interpersonal relationship, management people, team job, developing plans of action in continuous improvement, among others.

I am researching about of MAS receptor, belonging to the Angiotensin complex in ischemic murine models, these research probable will be my project for graduation. I have two institutions for applying, my first choice is the Free University of Berlin (fu-berlin) in Germany, which led me to this choice was, they have partnership for research with my advisor of graduation with similar research lines, with my future projects, and because I intend to live in Germany in the future, they are one of the best countries for researchers in the world.

The area of Understanding Neural Circuits, is my target for my career like research, how Neural Circuits function and how alterations in this circuits impact the behavior, in a cellular and molecular level. In fu-berlin they have the Menzel Group in Institute of Biology led by Prof. Randolph Menzel who is dedicated to study Neurobiology and the Behavior of the Honeybee.

My passage in this university will contribute for my grown like person and like with researcher. Always I dream study in another country, know new culture, news research centers, is one cluster of experiences and new challenges. Study abroad does not give a chance the access a news nucleus of research and new academics experiences, social and cultural experiences, I believed when I will be one cluster of unique experiences when will reflected positively in my career academic and professional.

TEXT 4

Dear professor, My name is XXX. I currently working on a PHD of arts at Universidade Federal de Minas Gerais (Brazil). I am searching the negative charge of the shadow in the myths, arts, and fine arts. I chose this field of study because I wanted to know why the shadow is the principal esthetic element of the German Expressionists Cinema mise en scène that can arouse fear and panic and what the shadow's functions in these films. I am doing this search with the help of my tutor PHD Luiz Nazario. We published papers of this subject in Revista Trama (Rio The Janeiro), in the Annals of the International Conference on Cinema, Art, Technology, Communication - Avanca 2017 (Portugal) and I will publish in the Annals of III Jornada de Cinema e Ficção Audiovisual (Curitiba/Brasil).

Therefore, these research of the shadow's functions led to me to the sounds' functions in the films. I know that in your prestigious University you are developing researches about these subject and, for this reason, I want to develop this part of my research in your prestigious institution. Due to their engagement of research on sound's functions in the films.

Thus, I will be so happy that you accept me in your University, and I promise that I will try to do the best of my research Taking advantage of all the resources that your institution can offer.

TEXT 5

I am a Brazilian, public employee, graduated in Information Science and Fashion Design student. At the moment I am doing the undergraduate in Fashion Design, at the UFMG (Universidade Federal de Minas Gerais). During this fashion course, here in Brazil, I have studied practices and theoretical subjects like modeling, tailoring, drawing of the human figure, handbags and textile technology.

I always had interest in drawing, making handicrafts like accessories and observing how the fashion influences the people's lives. I identify myself a bit with each ones, and I am also interested in teaching about those areas.

Before this undergraduated course I did free courses of draping and flat modeling. And after to conclude it I would like to unite and apply the Information Science with Fashion Design in my professional life.

Beyon to study fashion, in my day by day I have several others interests, like learning new languages, traveling, meeting new people and cultures, practicing sports, cooking and socializing with the friends.

As I know that Italy has big relevancy and influence in the fashion of the most countries around the world, I am sure that studying fashion there will value so much my curriculum. And to study at a renowned university such as the University of Bologna, in Milan - the fashion capital of the world - provide me to conclude my course with " golden key".

TEXT 6

When I was in the University Federal of Ceará I figured out why I like science and which makes me continue my studies after I graduated. During my graduate study, I cultivated a passion for plants only upon completion of an entry-level botany course and afterwards I was excited about joining in the laboratory of plant taxonomy. I then enrolled in additional courses in order to further focus on this botanical interest. To follow my passion for this area, I decided that I would like to improve my knowledge about plants then I studied abroad for one academic year at University of Pisa in Italy, in addition to the classes I have had while I was there I also did an internship for six months in a laboratory of botanic.

TEXT 7

My name is XXX and I am a student of Biomedicina at UFMG - Universidade Federal de Minas Gerais. I am 31 years old, I always lived in Belo Horizonte, a large city in the state of Minas Gerais - Brazil. I would to study in this university because my interests in Biomedical Science started when I was reading about DNA and Human Genoma Project, besides what I always liked [enjoyed] reading news and texts related to health, researchs on new methods of curing diseases and discoveries in topics related to biotechnology.

TEXT 8

My name is XXX, I am twenty two years old and I live in Minas Gerais in Brazil. I am studying Letters at Universidade Federal de Minas Gerais (UFMG), and I am in my second year. I decide on study Letters due to be a teacher its a important profession (...) During this time in Faculty of Letters, I begin to be very interesting in English education (dizer o pq). In the first semester of 2017, I worked for a project that was dedicated to students that aspire to be professors, to have their first experience in this job and learn about their area a little more. In this project we had the opportunity to study and see a little how was the process on a second language acquisition, and the experience to work as a teacher. I decide on California State University (CSU) – Chico whereas you have a College of communication and Education that has guides and missions such as (...) Furthermore your University is on the best universities of

TEXT 9

My name is XXX and I started my undergraduate studies at the Faculty of Economics from Federal University of Minas Gerais, in the field of Economics related to International relations. My main interest about my knowledge of Public Law and Economics field is to reconcile their differences in the best way possible.

My career aspirations are teachers and professors that I had during my High School and undergraduate time, who I had the pleasure to live with and to share academic experiences. I was fortunate to have a good intern and excellent professors during my academic life.

After two year's work experience as an intern on Public Ministry of Justice (on "Fazenda Pública") at my first graduation, I gain some practical experience working with Public Law and I could see how important the Economy was to Public Administration, especially macroeconomics and microeconomics, now I have interest in knowing more about public politics and how to make the World a better place to live through Economic planning.

My love of Economy started during my Post-Graduation Lato Sensu Course on Public Law field, there I met teachers who introduced to me a lot Economist authors, most of them with divergent perspectives from Economy.

From disciplines like "Fiscal Responsibility" and Financial Law I could understand how was important the economic planning to Public Administration at the same way that was important the correct application of the law. One professor also encouraged us to read Capital in the twenty-first century, a book from Piketty a very famous Economist, I remember clearly the impact reading these book had on me. This interdisciplinary aspect of the subject helped me make my choices.

The main characteristics of both professional areas are to be proactive, analytic, critical not only on professional time, but also during free time. I'm a person who likes to discover new things every time, I like to do research by myself about economics and law fields. I also like to read and to listen people who have different opinions from me about many varieties of things. I have interest in improve my knowledge about public politics and how to make the World a better place to live and at undergraduate level, my interests became more focused.

In selecting the University of Chicago I have been influenced by the way that this University treats their members, how they conduct the career planning and how profound influence of economic researches on a wide diversity of areas.

TEXT 10

My interest for electrical engineering began when I started a professional formation course in electrical industry at SENAI/FIEMG school, in 2008. After that, in first semester of 2009, excited to know more about these world, I decided enter in a electrical engineering program in high school, at Unileste university.

In 2009, in a third semester, I was selected to work in a academic project related with electrical installation of engineering campus, in Coronel Fabriciano, Minas Gerais. My activity in this project was, with other students, to mapping the campus electrical network infrastructure, also known as "as built". I worked in this project one year. After that, I had opportunity to work for two year with research in a academic program, that is called scientific initiation, supported by Minas Gerais fomentation agency, FAPEMIG. This experience was amazing, because I should to acquire knowledge in renewable energy, distributed generation and power quality. I was monitor of electrical machines discipline per three semester, developing practical activities at laboratory witch the professor and helping students. I also was a internship in two big companies at the still maker industry, Aperam South America, summer program, and Usiminas, trainee program. All these experience was fundamental to increase my knowledge and my career in electrical engineering. I conclude the bachelor degree in June 2014, and starter a Master program in electrical engineering at Federal University Of Minas Gerais - UFMG, in march 2015.

Nowadays, I am PhD. student at in electrical engineering at UFMG working in power electronics area with focus in power train for electric vehicles.

TEXT 11

On my career plan, I hope to work on Material Science as research science or make my own business. When I accumulate enough experience I will follow what I want. Absorbing new technologies is always a pleasure to me so I chose to study further. Studying abroad has been my dream since I was in junior high school. First of all, I wish to get an undergraduate on Materials Science and Engineering. While I study in the field of Materials, I hope to implement my ideas on industry in general, including finding methods more clean to produce. I chose the MIT because he is the best university in the world in this field.

When I begin on faculty I was coursing chemical engineering and on my first semester I got a scholarchip called "Jovens Talentos da Ciência" including an science project on physic that was supervised by Doctor Caio Olindo, my group research was "Interferometro de Fabry Perot com luz não monocromática", on the first months we just program the machine and made a bibliography review, together with the scholarship I got a chance to learn english with a program on the campus. Behind of student life I was aggregate of many.

TEXT 12

My name is XXX, I am 27 years old, I live in Belo Horizonte, Minas Gerais, Brazil. I am a Phd student in the Biochemistry Department from University of Minas Gerais, Brazil. I concluded my Master last year and graduated in Biology in 2013 at the same university.

In my graduation, I work with Medical Entomology in Renè Rachou, an institute of Fiocruz making cellular cultivate and take care of insectary. Then, I started work in Laboratory of Biochemical Genetics in UFMG, where I researched about DNA Repair in *Trypanosoma cruzi* and learned many techniques, such a cloning and production of recombinant proteins. Nowadays I work with Molecular Biology in Immunochemistry of Proteins and my research involves antivenoms. Dr. Carlos Chavez Olórtegui is specialist in antivenoms in Brazil and Peru, and I participate in his research group.

My work focus in loxoscelism, caused by Brown Spiders, or *Loxosceles* spp. The spiders of *Loxosceles* genus are very common in tropical areas and causes many accidents in the world. The victims present dermonecrosis and may have renal failure, leading to death. The most of antivenoms are produced by horses which receive the poisoning of the venom of interest. This animals suffer with several symptoms of envenoming and their life expectancy is reduced. Annually, thirty-six thousand spiders are used to obtain venom for this process. So, is very important search alternatives means of antivenoms production. In my project, we developed a non toxic recombinant protein who act with antigens and show great results, producing neutralizers antibody and maybe can substitute the venoms in sera production.

The Imperial College is a London science-based institution, consistently rated amongst the world's best universities. Our lab collaborates with Braga Lab, the lab of Dr. Vânia Braga, Imperial College member. We intend to strengthen our partnership with new experiments that I will conduct in her lab. Braga Lab have expertise in production of recombinant proteins in mammalian cells, and I would like learn this technique.

I have experience with immuno assays, neutralization assays and animals experiments, and I can help the colleagues of Braga Lab with this know how. It will be a pleasure to meet London and collaborate with this team.

TEXT 13

My life in the Sports environment began early, when I was around 4 years old, in the Soccer School, and later in the Jiu-Jitsu School, and since that time, I realized that I loved Sports, and I wanted spend my whole life with that. When I began in Undergraduate School, I already was sure that my work field will be in Sport Science. I get into Physical Education School at UFLA in 2014, after give up my first job as a Graphic Designer, which I worked for three years. However, I was not happy with the choice that I made, so, after receiving some advices from a friend, I searched deeply about the Physical Therapy are, then I completely fell in love!

So, in 2015 I started in Physical Therapy School at UFMG (rated as the best Physical Therapy program in Brazil) and always focusing in Sports area, searching and reading as much as I can, and getting in tutor programs for undergraduate students, I am threading my way through that universe. In second semester, I really figure out what I will go to study further: Exercise Physiology.

When I decided to search about the best Universities in that field, what I found was the University of Sydney, as the best in Science Sports, thus, I seek more and deeper about the School, then I am really hopeful about to accomplish my dream there, in the best University in the World! I could see there are so many alternatives, the place structure is awesome and so great researches and discoveries are made in Sydney, so I want to be a part of this history!

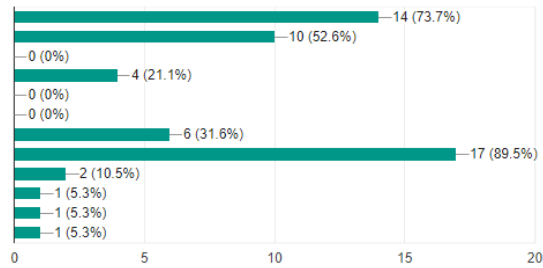
Moreover, the country seems perfect for me. The weather looks like the Brazilian climate, and everything that I heard about there, are compliments. A great people and a wealthy country with

many nature beauties. Hence, studying at University of Sydney, I will achieve two dreams: To meet that awesome place, and to study what I love in the best University in the World. I hope that someday I can share my expectations and to start a new life, with news perspectives and I really sure that is the best place for me.

APPENDIX 9 – Responses of the Second questionnaire

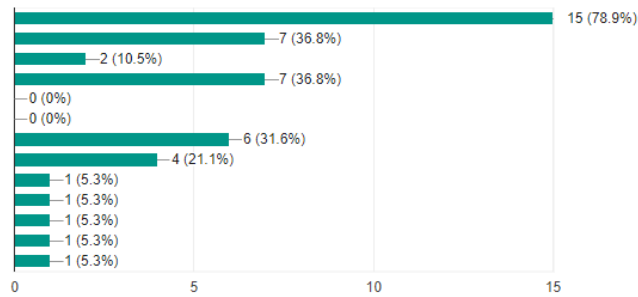
Quais ferramentas foram usadas por você para melhorar seu texto (segunda versão)?

19 responses



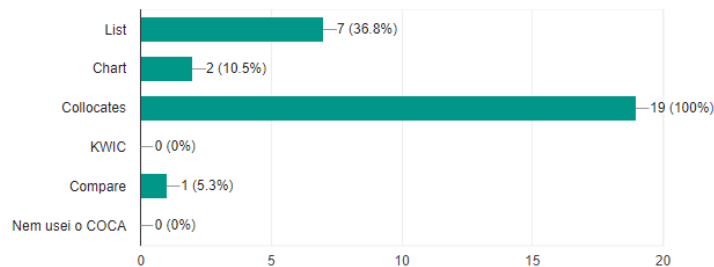
No dia-a-dia, ao escrever algo em inglês (e-mail, textos acadêmicos e outros), você utiliza...

19 responses



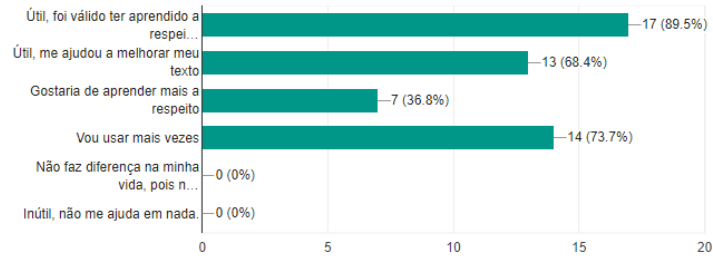
Quais foram as ferramentas do COCA mais usadas por você para corrigir seu SoP?

19 responses



A respeito do COCA, você considera...

19 responses



APPENDIX 10 – Tables with lexical and grammatical errors.

Table 12 - Lexical errors table

	Lexical errors	Student's Correction	GRADE
Sentence T1-1	My name is xxxx and a study english for a any years.	My name is xxxx, I am 38 years old and living in Belo Horizonte, at Minas Gerais State. I study english for 2 years.	A1
	My name is xxxx and a study english for a any years .	My name is xxxx, I am 38 years old and living in Belo Horizonte, at Minas Gerais State. I study english for 2 years .	A1
Sentence T1-2	I have so much difficulty with progres in english	I have so much difficulty with progres in English	A1
Sentence T1-3	I would lik to know try make a exchange with on university on USA or other english speaker country for improove my knowledge and my conversation.	But I hope to try one chance for to do an exchange.	B
Sentence T1-4	My research is about de middle of learning between observation and reproduce, indeed, my research is about "the copy".	My Scientific Initiation analysing the process of observation and reproducing, basically, I study the process of copy the movement through observation.	C
	My research is about de middle of learning between observation and reproduce, indeed, my research is about "the copy".	My Scientific Initiation analysing the process of observation and reproducing, basically, I study the process of copy the movement through observation.	A1
Sentence T2-2	But I have difficulty with English, and I intend to improve it by doing an exchange	I intend to improve it by doing an exchange	B
	But I have difficulty with English, and I intend to improve it by doing an exchange	I intend to improve it by doing an exchange	B
Sentence T3-5	I am participating in more than four projects in the Laboratory, helping my team in other research lines	I am participating in more than four projects, helping my team in other lines of research	A1
Sentence T3-6	I work since twelve years , because in city when i borne, they accept all the kids	I worked in a Company for communications during 7 years , where I get abilities for interpersonal relationship	A1
	I work since twelve years, because in city when i borne, they accept all the kids	I worked in a Company for communications during 7 years, where I get abilities for interpersonal relationship	A1
Sentence T4-2	We public papers of this subject in Revista Trama (Rio The Janeiro)	We published papers of this subject in Revista Trama (Rio The Janeiro)	A1

Sentence T4-3	I will public a paper in the GRUDES book of papers.	I will publish in the Annals of III Jornada de Cinema e Ficção Audiovisual (Curitiba/Brasil).	A1
Sentence T5-3	At the moment I am doing the undergraduate in fashion design	At the moment I am doing the undergraduate in Fashion Design	B
Sentence T6-2	During the time I was in the college, I applied to participate in a laboratory of sistematic of plants,	During my graduate study, I cultivated a passion for plants (...) and afterwards I was excited about joining in the laboratory of plant taxonomy.	A2
Sentence T7-2	During my graduate in Information Systems , I studied data bases xxx and now I am studying xxx	I am taking my second degree and during my graduate in Information Systems , I studied data structures and algorithms	B
Sentence T8-1	When I decided to course Letras, during my classes I started to being very excited about the studies in language acquisition.	I decide on study Letters (...) During this time in Faculty of Letters, I begin to be very interesting in English education	A1
	When I decided to course Letras , during my classes I started to being very excited about the studies in language acquisition.	I decide on study Letters (...) During this time in Faculty of Letters, I begin to be very interesting in English education	A1
Sentence T8-2	This interesting intensify because (·	-	C
Sentence T8-3	I started to work in a project that was dedicated to knew professor , to be their first jobs and know about their area a little more.	I worked in a project that was dedicated to students that aspire to be professors, to have their first experience in this job and learn about their area a little more.	A1
Sentence T9-1	I started my undergraduate studies at the Faculty of Economics from UFMG, in Economic field related to International relations.	I started my undergraduate studies at the Faculty of Economics from Federal University of Minas Gerais, in the field of Economics related to International relations.	A1
Sentence T9-2	My professional aspirations are teachers and professors that I had during my High School and undergraduate time	My career aspirations are teachers and professors that I had during my High School and undergraduate time	A1
Sentence T9-3	who I had the pleasure to live with and divide academic experiences .	who I had the pleasure to live with and to share academic experiences	A1
Sentence T9-4	I was fortunate to have good intern and excellent professors during my academic life.	I was fortunate to have a good intern and excellent professors during my academic life	B
Sentence T9-5	I gain some practical experience working with Public Law and I could see how important the economic was to Public Administration	I gain some practical experience working with Public Law and I could see how important the Economy was to Public Administration	A2
Sentence T9-6	now I have interest to know more about public politics and how to make a word a better place to live.	now I have interest in knowing more about public politics and how to make the World a better place to live through Economic planning.	A1
Sentence T9-7	My love of Economic started during my Post-Graduation Lato Sensu Course on Public Law field	My love of Economy started during my Post-Graduation Lato Sensu Course on Public Law field	A2
Sentence T9-8	I met teachers who introduce to me a lot Economics authors, most of them with divergent perspectives from economic	I met teachers who introduced to me a lot Economist authors, most of them with divergent perspectives from Economy .	A1
Sentence T11-3	First of all, I wish to get a undergraduate	First of all, I wish to get an undergraduate	B

Sentence T12-1	Actually , I work with Molecular Biology in Immunochemistry of Proteins and my research involves antivenoms.	Nowadays I work with Molecular Biology in Immunochemistry of Proteins and my research involves antivenoms.	A1
Sentence T12-2	Dr. Carlos Chavez Oliveira is specialist in antivenoms in Brazil e Peru, and I participate in your research group.	Dr. Carlos Chavez Oliveira is specialist in antivenoms in Brazil and Peru, and I participate in his research group.	A1
Sentence T13-2	In 2011, I went to my first job , as a Graphic Designer, and worked until 2014	I get into Physical Education School at UFLA in 2014, after give up my first job as a Graphic Designer, which I worked for three years.	C

The lexical errors were homophone induced errors, false cognates, orthographic errors, wrong words that in L2 (in this case, Portuguese) can be used in many situations but in English it does not apply.

Table 13 - Grammatical errors table

	Gramatical errors	Student's Correction	
Sentence T1-3	I would lik to know try make a exchange with on university on USA or other english speaker country for improve my knowledge and my conversation.	But I hope to try one chance for to do an exchange.	A2
	I would lik to know try make a exchange with on university on USA or other english speaker country for improve my knowledge and my conversation.	But I hope to try one chance for to do an exchange .	A1
	I would lik to know try make a exchange with on university on USA or other english speaker country for improve my knowledge and my conversation.		C
	I would lik to know try make a exchange with on university on USA or other english speaker country for improve my knowledge and my conversation.	These process help me to improve the knowledge on my field research	A1
Sentence T1-5	<u>others researchers</u>	other directions	A1
Sentence T2-1	I'm student of Civil Engineering on UFMG ,	I'm student of Civil Engineering in UFMG	A2
Sentence T2-3	I would like know Australia	I would like go to University of Sydney	A2
Sentence T3-1	I am single, i borne in Porteirinha	I am single, I was born in Porteirinha	A1
Sentence T3-2	I am loved Biology	I always dreamed of being a biologist	A1

Sentence T3-3	my interest in Biology is in Neuroscience, the part of physiology, how happens interactions between the neurons.	my interest area is Neuroscience.	C
Sentence T3-4	i dream go to Canada or Germany, and continue to studs	My dream is to study in Germany , and continue to study	A1
	i dream go to Canada or Germany, and continue to studs	My dream is to study in Germany, and continue to study	A1
Sentence T3-5	i am research about influence of MAS receptor	I am researching about of MAS receptor	A1
Sentence T3-6	I work since twelve years, because in city when i borne, they accept all the kids	I worked in a Company for communications during 7 years, where I get abilities for interpersonal relationship	A2
	I work since twelve years , because in city when i borne, they accept all the kids	I worked in a Company for communications during 7 years , where I get abilities for interpersonal relationship	A1
	I work since twelve years, because in city when i borne , they accept all the kids	I worked in a Company for communications during 7 years, where I get abilities for interpersonal relationship	C
Sentence T4-1	I want to know why the shadow is the principal esthetic element of the fine arts mise en sc�ne to can arouse fear and panic.	I wanted to know why the shadow is the principal esthetic element of the German Expressionits Cinema mise en sc�ne that can arouse fear and panic	A1
Sentence T5-1	I am always have interest in draw	I always had interest in drawing	A2
	I am always have interest in draw	I always had interest in drawing	A1
Sentence T5-2	how the fashion influences the people life	how the fashion influences the people's lives .	A1
	how the fashion influences the people life	how the fashion influences the people's lives .	A1
Sentence T5-4	I would like to unite and apply Information Science and Fashion Design in my professional life .	I would like to unite and apply the Information Science with Fashion Design in my professional life .	A2
Sentence T6-1	In the college I figured out why I like science and what makes me to keep continue my studies after I graduated.	When I was in the University Federal of Cear • I figured out why I like science and which makes me continue my studies after I graduated.	A2
	In the college I figured out why I like science and what makes me to keep continue my studies after I graduated.	When I was in the University Federal of Cear • I figured out why I like science and which makes me continue my studies after I graduated.	A1
Sentence T6-2	During the time I was in the college , I applied to participate in a laboratory of sistematic of plants,	During my graduate study , I cultivated a passion for plants (...) and afterwards I was excited about joining in the laboratory of plant taxonomy.	A1
Sentence T7-1	I would to study in this university	I would to study in this university	B

Sentence T8-1	When I decided to course Letras, during my classes I started to being very excited about the studies in language acquisition.	I decide on study Letters (...) During this time in Faculty of Letters, I begin to be very interesting in English education	A1
Sentence T8-2	This interesting intensify because (•	-	C
Sentence T8-4	last semester I toked class of Oral Expression that was a discipline that to help us with our English skills	-	C
	last semester I toked class of Oral Expression that was a discipline that to help us with our English skills	-	C
Sentence T8-5	we had to learn about how was the processes of listening a different language, based in • , to help us to discuss about this topic in the classes.	-	C
	we had to learn about how was the processes of listening a different language , based in • , to help us to discuss about this topic in the classes.	-	C
	we had to learn about how was the processes of listening a different language, based in • , to help us to discuss about this topic in the classes.	-	C
Sentence T8-6	This help me to to decided CHICO university because you have a group related of education.	I decide on California State University (CSU) – Chico	A1
Sentence T9-6	now I have interest to know more about public politics and how to make a word a better place to live.	now I have interest in knowing more about public politics and how to make the World a better place to live through Economic planning.	A1
Sentence T9-7	My love of Economic started during my Post-Graduation Lato Sensu Course on Public Law field	My love of Economy started during my Post-Graduation Lato Sensu Course on Public Law field	A2
Sentence T9-8	I met teachers who introduce to me a lot Economics authors, most of them with divergent perspectives from economic	I met teachers who introduced to me a lot Economist authors, most of them with divergent perspectives from Economy.	A1
Sentence T9-9	This interdisciplinary aspect of the subject helped me to make my choices.	This interdisciplinary aspect of the subject helped me make my choices.	A1
Sentence T10-1	After that, in first semester of 2009,	After that, in first semester of 2009	B
Sentence T11-1	i hope to work on Material Science as research science or make my own business.	I hope to work on Material Science as research science or make my own business.	B
Sentence T11-2	I will try follow what i want	I will follow what I want	A1
Sentence T11-3	First of all, I wish to get a undergraduate	First of all, I wish to get an undergraduate	A1

Sentence T11-4	I hope to implement my ideas on industry in general, including find methods more clean to produce.	I hope to implement my ideas on industry in general, including finding methods more clean to produce.	A1
	I hope to implement my ideas on industry in general, including find methods more clean to produce.	I hope to implement my ideas on industry in general, including finding methods more clean to produce.	B
Sentence T13-1	and always focusing in that area	and always focusing in Sports area	B

The grammatical errors were mostly agreement errors, tense, singular-plural, prepositional errors, article and conjunctions errors.