UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

INSTITUTO DE LETRAS

MARINE LAÍSA MATTE

A CORPUS-BASED STUDY OF CONNECTORS IN STUDENT ACADEMIC WRITING

Porto Alegre

2017

MARINE LAÍSA MATTE

A CORPUS-BASED STUDY OF CONNECTORS IN STUDENT ACADEMIC WRITING

Trabalho de conclusão de curso de graduação apresentado como requisito parcial para a obtenção do grau de Licenciada em Letras pela Universidade Federal do Rio Grande do Sul.

Orientadora: Profª. Drª. Simone Sarmento

Porto Alegre

2017
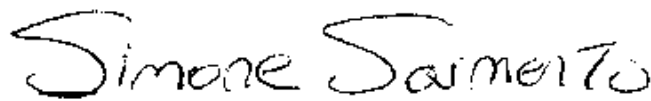
MARINE LAÍSA MATTE

A CORPUS-BASED STUDY OF CONNECTORS IN STUDENT ACADEMIC WRITING

Trabalho de conclusão de curso de graduação apresentado como requisito parcial para a obtenção do grau de Licenciada em Letras pela Universidade Federal do Rio Grande do Sul.

Orientadora: Profª. Drª. Simone Sarmento

BANCA EXAMINADORA

_____

Simone Sarmento

Universidade Federal do Rio Grande do Sul

_____

Larissa Goulart da Silva

Universidade de Warwick

_____

Rozane Rodrigues Rebechi

ACKNOWLEDGMENTS

ABSTRACT
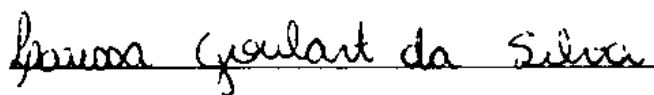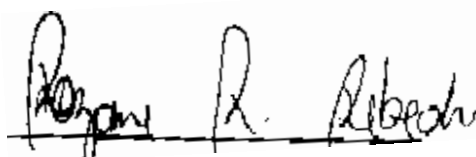
This monograph is about the use of connectors in a corpus composed of assignments written by Brazilians studying in British universities (BrAWE) as compared to another corpus which comprises academic texts written by a group of students with high marks in British universities (BAWE). Connectors are important because they are one of the linguistic devices that provide cohesion to the text. This paper brings together theoretical assumptions from Corpus Linguistics, English for Academic Purposes and connectors. The results indicate that there is a statistically significant difference in the use of the entire range of connectors in the investigated corpora, with Brazilian students overusing connectors as a whole and, more specifically, connectors which express additional ideas. This research points to the importance of corpus-based studies as they provide evidences of language production and help in the creation of pedagogical material for academic English classrooms.

Keywords: Corpus Linguistics; English for Academic Purposes; Connectors.

RESUMO

Este trabalho trata dos usos de conectores em um corpus composto por trabalhos escritos por estudantes brasileiros realizando seus estudos em universidades da Inglaterra (BrAWE). O objetivo é analisar de que forma esses usos se diferenciam daqueles realizados por um grupo de estudantes que receberam notas elevadas em universidades britânicas (BAWE). Conectores são importantes porque atuam como um dos mecanismos linguísticos que conferem coesão ao texto. O trabalho conta com referenciais teóricos da Linguística de Corpus, de Inglês para Fins Acadêmicos e de conectores. Os resultados indicam que há uma diferença significativa nos usos dos conectores nos corpora estudados, bem como uma tendência de uso excessivo desses elementos linguísticos que expressam ideias de adição por parte dos estudantes brasileiros. Essa pesquisa revela a importância de estudos baseados em corpus, uma vez que se tem evidências de produção da língua, o que auxilia na criação de materiais pedagógicos para a sala de aula de inglês acadêmico.

Palavras-chave: Linguística de Corpus; Inglês para Fins Acadêmicos; Conectores.

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

BAWE - British Academic Written English

BrAWE - Brazilian Academic Written English

CL - Corpus Linguistics

EAP - English for Academic Purposes

EFL - English as a Foreign Language

EGAP - English for General Academic Purposes

EGP - English for General Purposes

ELT - English Language Teaching

ESAP - English for Specific Academic Purposes

ESP - English for Specific Purposes

GE – General English

LL – Log-Likelihood

LwB - Languages without Borders

SwB – Science without Borders

TABLE OF CONTENTS

1. INTRODUCTION

When learning an additional language, it is every learner's wish to achieve a certain autonomy when it comes to reading, writing and understanding texts. And these three actions are possible if the focus is, among others, on the vocabulary used in these texts. Following the argument of Nation (2013, p. 4), "knowing a word is taken to include not only knowing the formal aspects of the word and knowing its meaning, but also being able to use the word", it is worth mentioning the importance of dealing with vocabulary in context. In other words, it is nonsense to learn isolated words if the student does not know how to use them for real purposes.

Regarding academic English, the approach for learning vocabulary is not different. Coxhead (2011) claims that knowing a considerable amount of academic words in a text is determinant for the student to get an almost complete comprehension of the text. Moreover, it is known that native speakers of English have a different performance when using the language to write if compared to non-native users for whom English is an additional language. (BERMAN; CHENG, 2010).

There are some words and expressions, such as connectors[1], that are essential to all users wishing to master their writing skills (PLANCIC; NINCEVIC, 2014). Even though they are grammar words that do not fit into the classification of general academic vocabulary,[2] I believe it is essential to establish that these lexical items are as important as knowing academic words in order to write successfully. Halliday and Hassan (1976, p. 226) claim that connectors are cohesive devices since "they express certain meanings which presuppose the presence of other components in the discourse". With this in mind, the intention of this paper is to analyze the uses of connectors in Brazilian written assignments, due to the fact that the cohesion of a text depends on the adequate use of these linguistic elements. To this end, this study uses corpus linguistics as a framework to investigate how Brazilian students use connectors when writing academic English.

---

[1] Connectors are used to join separate sentences. By using a connector, the separation between two statements allows their relationship to be more precisely defined, which is especially important for professional writing. Therefore, connectors relate to the previous clause or sentences, and it often goes in front position. (EASTWOOD, 1994)

[2] Academic vocabulary refers to words used in texts across several disciplines in academic contexts. Therefore, these words are different from those used in our daily lives. (BAUMAN; GRAVEZ, 2010; CLARK. ISHIDA, 2005)

This chapter introduces the study as well as the motivation to pursue the topic of connectors (1.1). In the second section (1.2), I explain the aims of the study and what research questions guide it. Finally, in the last section (1.3) I present how the study is organized.

1.1 Motivation

The motivation to conduct a study on written academic English derives from my experience as an English teacher at the Languages without Borders (LwB) Program, which aims at promoting the internationalization of Brazilian higher education through the provision of English classes and administration of English language proficiency exams. LwB aims at undergraduate and graduate students, professors and administrative staff of federal universities. The LwB program offers a variety of courses with different aims and lengths; for instance, the courses can focus on general English (English for General Purposes - EGP), specific areas (English for Specific Purposes - ESP) or English for Academic Purposes (EAP). A great amount of research has been conducted, from the analysis of the program itself (SARMENTO; ABREU-E-LIMA; MORAES FILHO, 2016) to detailed descriptions of how coursebooks are used in the classroom. (LAMBERTS, 2015; VIAL, 2014)

According to Welp, Sarmento and Fontes[3] (2016, p. 141-142),

> [T]he Languages without Borders program have contributed to many areas involved in the process of internacionalization of Brazilian universities, in teacher training and in language teaching (…). These different fronts foster a rich scenario for knowledge production. Therefore, numerous studies are emerging from higher education institutions involved in the program.

Regarding my experience as an English teacher and having to deal with students from all levels of proficiency, I have realized how difficult it is to develop their arguments through a consistent use of connectors. Due to the fact that they are Brazilians, they probably tend to use connectors that are similar to Portuguese[4], but this does not guarantee a writing of quality. On the contrary, this might reveal a lack of

---

[3] Translation made by the author. Original quote: "O Programa Idiomas sem Fronteiras veio contribuir para muitas áreas envolvidas no processo de internacionalização das universidades brasileiras, da formação de professores e do ensino de línguas ao estudo de políticas públicas. As diversas frentes implicadas propiciam um cenário farto para a produção de conhecimento. Nesse sentido, são inúmeras as pesquisas que estão surgindo nas instituições de ensino superior envolvidas no programa." (WELP, SARMENTO e FONTES, 2016, p. 141-142)

[4] By this, we mean connectors similar in form like 'in addition'.

familiarity with the most common words that are used to connect ideas. Hence, connectors are the object of this study.

The next section (1.2) presents the aims of the study and the research questions that guide it.

1.2 Aims of the study

First, the major objective of this study is to unveil Brazilian students' use of connectors and compare them to students represented in a British corpus. Finding out about the difference is important because connectors are difficult to master despite proficiency level, and they are related to cohesion, i.e., intelligibility. Additionally, "increased mastery of cohesive devices will help students to express relations more clearly." (GRANGER; TYSON, 1996, p. 26). Hence, the data are extracted from two academic corpora of written English: the British Academic Written English (BAWE) and the Brazilian Academic Written English (BrAWE). These corpora are going to be detailed in Chapter 3. Moreover, the findings of this study will contribute to English teachers who will be able to make more sound decisions about what to include in their lesson plans in terms of connectors.

Several studies have been conducted to analyze the uses of connectors by English language learners with different mother tongues (GRANGER; TYSON, 1996; MILTON, 2001; NARITA; SATO; SUGIURA, 2004; TAPPER, 2005), but, apparently, none has been found in terms of how Brazilians use connectors in English. Therefore, this study aims at answering the following questions:

a. How do Brazilian students in the UK use connectors in their written assignments in English?

b. Do these uses differ from students represented in the British Academic Written English (BAWE) corpus?

1.3 Organization of this study

This study is divided into five chapters. In chapter 2, a literature review is conducted related to Academic English and to how connectors are important cohesion devices and present a challenge for learners. Chapter 3 details the methodological

approach used to collect and analyze the data. The results and findings are shown in chapter 4. Finally, chapter 5 is aimed at developing some conclusions, which are also related to the pedagogical implications of this study.

2. LITERATURE REVIEW

This chapter addresses the topic of Academic English (2.1). The following section (2.2) gives an explanation on connectors, presents different studies on English learners use of connectors in English (2.2.1) and describes their importance for text cohesion (2.2.2)

2.1 English for Academic Purposes (EAP)

Although the expression EAP is frequently heard, some might not have a clear understanding of what this concept means. Emerging from the field of English for Specific Purposes (ESP) (see figure 1), EAP "refers to language research and instruction that focuses on the specific communicative needs and practices of particular groups in academic contexts." (HYLAND and HAMP-LYONS, 2002, p. 2). Based on this definition, an EAP classroom has the academic community as the target audience, as in the case of some LwB program courses whose main objective is to teach academic English. Considering academic English students, they look for a class that can improve their knowledge of academic uses of the language. Clark and Ishida (2005, p. 227) state that "[s]ince EAP learners are those with academic goals, it is natural that they need to be able to read academic texts well in their fields of study in order to achieve their goals." Thus, the teaching of EAP is for those students whose needs of using the language are related to uses in academic contexts. That is, based on the demands of a certain area, the students can look for academic English classes to learn how to write a specific genre (GARDNER; NESI, 2013), to learn more academic vocabulary or also to be exposed to technical vocabularies related to their own area of expertise.

The figure below is an attempt to differentiate EAP from English for General Purposes (EGP):

Figure 1 - ELT Tree
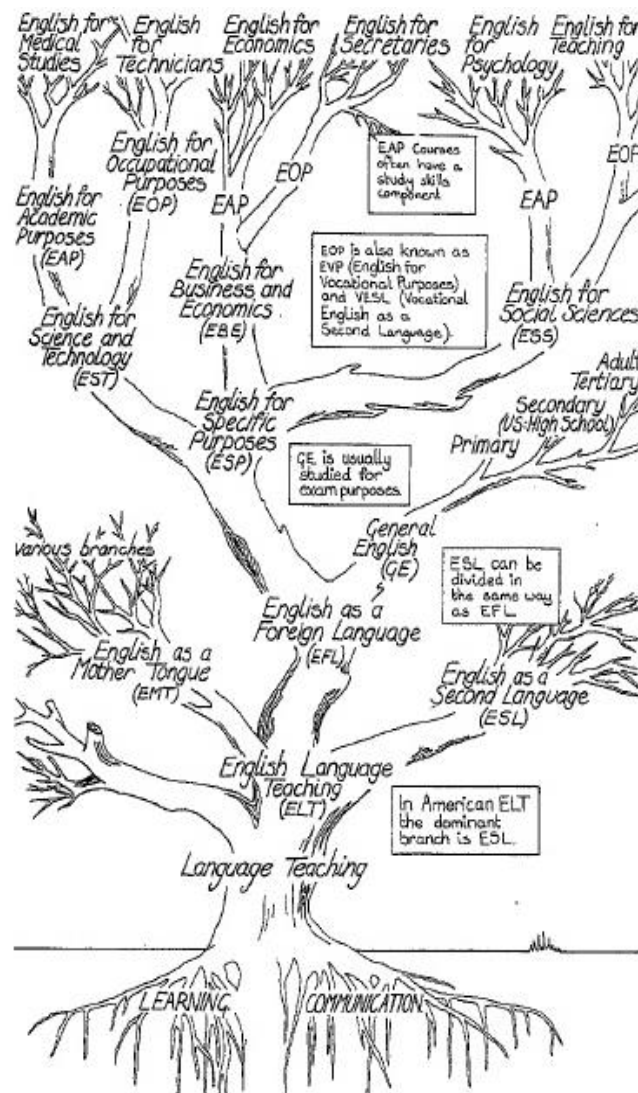
Source: Hutchinson; Waters, 1987

From the picture above, it is possible to observe that learning and communication are the two main roots of the Language Teaching tree. In this case, different branches emerge from an English Language Teaching tree. EAP derives from ESP, which, in turns, derives from English as a Foreign Language (EFL), diverging from General English (GE). This detour stems from the fact that EAP, and, hence, ESP, have particularities and different purposes when it comes to the English classroom.

Hyland (2016, p. 17) clarifies the differences between EGP and EAP by stating that "[w]hat sets English for Academic Purposes (EAP) apart from general language study is its focus on specific, purposeful uses of language." Based on this idea, he brings up some specificities in the EAP field, since there is the general EAP and the specific EAP, what he calls English for General Academic Purposes (EGAP) and

English for Specific Academic Purposes (ESAP). The focus here is only on the general approach of EAP as this study deals only with academic corpora, not taking into consideration specificities of different areas of expertise represented in both BAWE and BrAWE.

In EGAP, teachers are concerned with what is common to all areas in terms of academic discourse. In this sense, in EGAP lessons students learn the language to use in academic contexts, such as participating in conferences, taking notes on lectures or making oral presentations. These practices are universally useful to all students of different areas of expertise.

2.2 Connectors

Resorting to Longman Grammar of Spoken and Written English (LSWE), *connectors* are not a grammatical group of words. Instead, particular words have the function of connecting ideas, such as conjunctions and linking adverbials. Before going deep into the topic, it is worth mentioning the subtle difference between conjunctions and linking adverbials. At a first glance, they seem synonymous, since both of them join ideas. Nevertheless, they do not present the same linguistic characteristics, because linking adverbials are used to join ideas in two separate sentences or paragraphs, whereas conjunctions join ideas in the same sentence. For instance, in the sentence *I like to go for a walk because it keeps me fit*, 'because' is a conjunction. In "*I play soccer. However, I am not a huge fan of watching soccer on TV*," 'however' is a linking adverbial because it joins two ideas in separate sentences. The focus of this study is the analysis of linking adverbials only, because the interest here is on how Brazilian students join separate sentences, that is, how they concatenate their arguments throughout a text.

Figure 2 - Connectors as an umbrella term
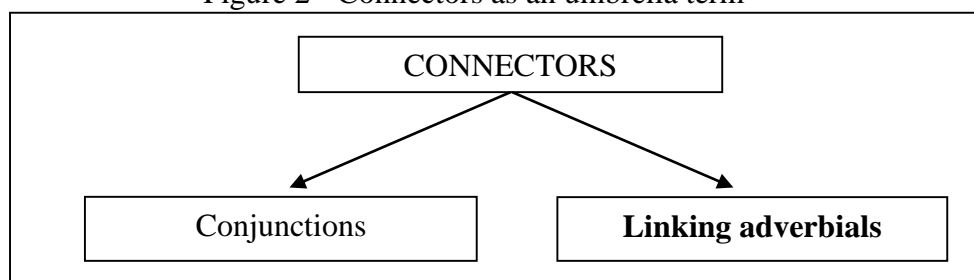
Source: the author

According to Biber et al. (1999, p. 875),

> the primary function of linking adverbials is to state the speaker/writer's perception of the relationship between two units of discourse. Because they explicitly signal the connection between passages of text, linking adverbials are important devices for creating textual cohesion, alongside coordinators and subordinators.

The passage above corroborates the idea that linking adverbials undertake a connective function and "[they] can express a variety of relationships, including addition and enumeration, summation, apposition, result/inference, contrast/concession, and transition." (BIBER et al, 1999, p. 765). Table 1 below presents some examples for each category mentioned by Biber et al. (1999)

Table 1: Linking adverbials categories (BIBER et al, 1999)

Category	Examples
Enumeration and addition	First, second, finally, lastly, furthermore, moreover
Summation	In sum, to conclude, overall, to summarize
Apposition	In other words, that is, for example, for instance
Result/Inference	Therefore, consequently, thus, so, then
Contrast/Concession	On the other hand, in contrast, though, however, alternatively
Transition	Incidentally, by the by, by the way

Source: the author

According to the LSWE, linking adverbials are most common in academic prose, since "a very important aspect [of this genre] is presenting and supporting arguments" (1999, p. 880). For this reason, these linguistic elements are frequent in written assignments. Considering the LwB program context, the students are asked to produce a considerable amount of written tasks, in which, and this is my perception, they develop their arguments not always knowing how to link the ideas and how to sustain their points of view with proper discourse elements.

2.2.1 The use of connectors: a challenge for EFL users

Catalán and Alba (2014, p. 199) conducted a study in which the objective was to make teachers and researchers aware of the "difficulties that foreign language learners face as far as the use of English connectors is concerned." Analyzing written essays of Spanish intermediate learners of English, the authors found three main types of problems with the use of connectors: total omission and scarcity of connectors (omission); connectors used unnecessarily (underuse); wrong choice of connectors (misused). The same issues were observed in previous studies, which are going to be detailed in the following lines.

Granger and Tyson (1996) put their effort on an analysis of French learners of English using connectors, and reached the conclusion that these learners tend to overuse connectors that perform additional ('moreover') and appositional ('for example') functions. Additionally, their study shows an underuse of connectors that "change the direction of the argument or take the argument logically forward." (p. 20), such as 'however' and 'therefore'. These learners seem to have misused them as an interference of their mother tongue. Granger and Tyson (1996, p. 22) also point out that another reason for learners to over- or underuse these discourse elements is the "lack of detailed description of connector usage in the dictionaries."

Milton (2001) focused his analysis on the uses of connectors of Chinese learners of English. By comparing their written productions with British students' written data, and by using a computational and corpus-based methodology, he observed cases of both over- and underuse of connectors. Just as Granger and Tyson's data, Chinese learners of English tend to overuse the linking adverbial 'moreover', as well as to use a large number of 'furthermore' and 'besides'. Regarding underused connectors, Milton (2001) lists the following ones: 'however', 'although' and 'for example'.

Japanese learners of English were also object of study in terms of the way they use connectors in English. Based on a quantitative analysis of connectors in advanced Japanese students' written texts in English, Narita, Sato and Sugiura (2004) compared these essay writings with the essays written by native English speakers. Their findings concern over- and underused connectors as well as the sentence position in which the connectors most appear. Additional connectors tend to be overused, such as 'moreover' and 'in addition'. Regarding the underused connectors, the authors mention 'however' and 'then'. It is worth mentioning that this study tries to explain the cases of underuse of contrastive connectors by claiming that "EFL learners are less familiar with the usage of

these rather formal contrastive connectors and thus they are likely to use other semantic equivalents that are already familiar to them in order to provide contrastive information." (NARITA; SATO; SUGIURA, 2004, p. 1174)

Regarding the uses of connectors made by Swedish advanced learners of English, Tapper (2005) compared these students' written English to American University students. The data were collected from the Swedish and American sub-corpora of the International Corpus of Learner Language (ICLE). The overall conclusion is that Swedish learners tend to overuse connectors. Moreover, when comparing to American students' uses, the author observes that written English by the Swedish contains more types of connectors. According to the connector classification Tapper (2005) uses, Swedish learners overuse clarifying connectors, such as 'that is', 'in other words', 'for example'. These words are what Biber et al (1999) call appositive linking adverbials. Furthermore, Tapper (2005, p. 124) points out that "only slight differences in the uses of the semantic functions by the Swedish learners and American students in the ICLE corpus were found" which is caused probably due to similarities between Swedish and English.

The table below systematizes the most common connectors used by learners of EFL in terms of over and underuse:

Table 2: Over - and underuse of connectors by learners of EFL

	Overuse	**Underuse**
Granger and Tyson (1996) (French)	Moreover, for example	However, therefore
Milton (2001) (Chinese)	Moreover, furthermore, besides	However, although, for example
Narita, Sato and Sugiura (2004) (Japanese)	Moreover, for example, in addition	However, then
Tapper (2005) (Swedish)	For example, in other words	------

Source: the author

This table suggests that even advanced learners overuse connectors which perform an additional function within the text. On the other hand, these data reveal that when it comes to conveying contrastive and conclusive ideas, leaners underuse connectors such as 'however' and 'therefore'. The relation between language use and proficiency was taken into account in Narita, Sato and Sugiura's study (2004). However, in the present study, proficiency level was not considered.

Considering that the main purpose of this study is to analyze the way Brazilians use connectors when writing in English, the aim of this chapter was to detail some studies regarding the uses of connectors by learners from different nationalities. Thus, it is possible to observe that in fact non-native English speakers present some degree of difficulties when it comes to using these linguistic elements in their written productions. Therefore, English teachers should have a special focus on connectors so as to ensure a higher level of cohesion for the text. This topic is going to be addressed in the next section.

2.2.2 Connectors as cohesion devices

Considering that connectors are used in a text, whether be it written or spoken, it is first necessary to specify what 'text' means. Defined as a basic element of human communication, a text is a "sequence of connections among various elements: sounds, words, meanings, discourse, participants, actions in a plan, and so on." (BEAUGRANDE, 1997, p. 11). This definition leads us to an understanding that a text is a whole of meaning, that symbolizes something, and that communicates something. (BARROS, 1990). Moreover, Barros (1990, p. 12)[5] claims that "a text only exists when conceived in its duality that defines it - an object of significance and an object of communication". Thus, every text is produced with a certain purpose, the reason why it needs to be intelligible in order to convey the intended meaning. In other words, cohesion is what guarantees the accuracy in meaning.

When Halliday and Hassan (1976) developed the notion of cohesion in English, they claimed that it is expressed in a text, which in turn is a unified whole. Besides, the authors also state that "[a] text is a unit of language in use" that can be "best regarded as a semantic unit: a unit not of form but of meaning." (1976, p. 2). By being a semantic unit, the text must contain elements that help them create this whole of meaning, and cohesion is one example of what the authors understand as semantic elements.

Hence, the concept of cohesion was used as a basis for this investigation. Considering the fact that the objective here is to analyze connectors in English from Brazilian students, these assignments are considered texts that must contain a certain level of cohesion in order to be intelligible. Halliday and Hassan (1976, p. 4) explain the

---

[5] Translation made by the author. Original quote: "[...]o texto só existe quando concebido na dualidade que o define — objeto de significação e objeto de comunicação [...]." (BARROS, 1990, p. 12).

way different linguistic elements build the cohesion of texts in English and assert that cohesion is an essential element to guarantee a writing of quality:

> [c]ohesion occurs where the interpretation of some element in the discourse is dependent on that of another. The one presupposes the other in the sense that it cannot be effectively decoded except by recourse to it. When this happens, a re1ation of cohesion is set up, and the two elements, the presupposing and the presupposed are thereby at least potentially integrated into a text.

Moreover, they consider cohesion as "a semantic relation between an element in the text and some other element that is crucial to the interpretation of it." (HALLIDAY HASSAN, 1976, p. 8). Therefore, if cohesion is the relation established by elements in the text, connectors work as the bridge between them, in the sense that they are linking words that play the crucial role of connecting ideas.

By referring to the authors above, Goldman and Murray (1992) conducted a study in which they analyzed the use of connectors and how determined they are for the comprehension of a text. By comparing the perceptions of connector's uses of both native and non-native speakers of English, they concluded that incorrect choices of connectors are related to the level of proficiency in English, which tends to be the case for learners of EFL. According to them,

> [c]onnectors, or conjunctives, are one class of signals that seem to be particularly important for expository text comprehension. Connectors are a type of text device the primary function of which is to link adjacent text propositions in such a way as to maintain local coherence." (GOLDMAN; MURRAY, 1992, p. 504)

In addition, they also see the relevance of connectors as elements that guarantee the cohesion of a text.

The next chapter (3) presents the methodological approach used in this investigation (3.1) by describing the academic corpora from which the data were extracted - BAWE (3.2) and BrAWE (3.3). The last section (3.4) contains the steps taken in the analysis.

3. METHODOLOGY

As previously mentioned, this study uses corpus linguistics to investigate how Brazilian students use connectors in their written assignments across British universities. In this chapter, I will present corpus linguistics as the methodological approach (3.1) and describe the corpora used for the analysis (3.2 and 3.3). In addition, in the last section (3.4) I will explain the steps taken to analyze the corpus.

3.1 Corpus linguistics

This investigation adopts corpus linguistics as its methodology. McEnery and Hardie (2011, p.1) point out that "the development of corpus linguistics has […] spawned, or at least facilitated the exploration of, new theories of language – theories which draw their inspiration from attested language use and the findings drawn from it."

From the authors' view, corpus linguistics is a resource to help deal with the language, and whose compilation of texts allows the researcher to explore uses of the language in real contexts. For Sardinha (2000, p. 325)[6], this linguistic approach has to do with "collecting and exploring corpora. […] Consequently, it focuses on language exploitation through empirical evidences, extracted from a computer."

In addition, regarding the possibilities of using academic corpora to develop research, Flowerdew (2004, p. 12-13) claims that "[c]orpus analysis provide attested examples of recurring language patterns, which are based on empirical data rather than introspection or gathered through elicitation techniques."

Corpora should contain authentic texts, that is, texts that have a communicative purpose and are not produced just to fit the corpora. As argued by Sarmento (2010, p. 100-101)[7], some of the great benefits of using a corpus approach is that we can "[…] extract typical and authentic examples of uses of a certain lexical item from a great amount of data in just a few seconds."

---

[6] Translation made by the author. Original quote: "A Linguística de Corpus ocupa-se da coleta e exploração de corpora, ou conjuntos de dados linguísticos textuais que foram coletados criteriosamente com o propósito de servirem para a pesquisa de uma língua ou variedade linguística. Como tal, dedica-se à exploração da linguagem através de evidências empíricas, extraídas por meio de computador." (SARDINHA, 2000, p. 325)

[7] Translation made by the author. Original quote: "A maior vantagem do uso de corpora na lexicografia é de natureza automatizada que permite que lexicógrafos consigam extrair exemplos típicos e autênticos do uso de um item lexical de uma grande quantidade de dados em apenas alguns segundos." (SARMENTO, 2010, p. 100-101)

This investigation uses two academic English corpora to analyze the uses of connectors. In the two following sections (3.2) and (3.3), I will give a brief description of both.

3.2 BAWE: British Academic Written English

The BAWE corpus was developed by Hilary Nesi and Sheena Gardner (Warwick), Paul Thompson (Reading), and Paul Wickens (Oxford Brookes) from 2004 to 2007. It contains assignments collected at Warwick University, Reading University, Oxford Brookes University, and later a few at Coventry University, representing written assignments of students across UK universities regardless of their nationalities. These assignments were all written in English and were submitted electronically. Moreover, the students' subject tutors judged them as standard productions in their areas. (ALSOP; NESI, 2009). Therefore, they were all merit and distinction assignments.

Regarding the corpus size of tokens, BAWE can be considered a medium-large-size corpus. According to Sardinha (2000), a corpus with 1 million to 10 million tokens is medium-large.

Table 3: Words and tokens in BAWE

BAWE	Words[8]	Tokens[9]
	6 968 089	8 336 262

Source: the author

Alsop and Nesi (2013, p. 72) state that "[t]he BAWE corpus is intended to enable the identification and description of student writing genres across disciplines and at different stages of academic development." The disciplines represented in the corpus are Life Sciences (LS), Social Sciences (SS), Arts and Humanities (AH), and Physical Sciences (PS). Although it is alleged a balanced corpus, it contains more texts in AH and SS if compared to PS and LS productions. According to Silva (2016, p. 42), these texts contain some meta data "such as students' level of education, their grades, previous study background, gender, among other information." For this reason, this

---

[8] In table 3 and table 4 (p. 24, 25), it is possible to observe that there is a quantitative difference between words and tokens because Sketch Engine tools count punctuation marks (commas, periods, semicolon) as tokens. Thus, the number of tokens is larger than the total of words.

[9] Tokens are words separated by spaces or punctuation, including repetitions of the same words. In *I study English and I love this language*, there are eight tokens and seven types, because types are word forms whose repetitions are not considered.

corpus has more detailed information when compared to the other corpus used in this investigation.

3.3 BrAWE: Brazilian Academic Written English

The BrAWE corpus was compiled by Silva (2016) as a result of her master's project at Warwick University in England. It is a medium-size corpus (SARDINHA, 2000), as it contains around 500,000 tokens (see table 4). In order for it to be compiled, the researcher gathered 380 written assignments of Brazilian students participating in undergraduate or masters programs in the UK. Altogether, there are 186 students from 59 universities represented in the corpus.

Table 4: Words and tokens in BrAWE

BrAWE	Words	Tokens
	657 859	768 323

Source: the author

Following the same categorization of BAWE, BrAWE has written productions from four main areas - Social Sciences (SS), Arts and Humanities (AH), Physical Sciences (PS), and Life Sciences (LS). Apart from the difference in size, there is another aspect that distinguishes both corpus, which is, "[…] contrary to BAWE which contained more texts in AH and SS than in PS and LS (ALSOP; NESI, 2009), the AH and the SS partitions of the corpus presented here are significantly smaller than the other two […]." (SILVA, 2016, p. 42) This corpus was compiled during the SwB[10] program and most of the students represented in it were SwB[11] scholars. Therefore, the scholarships were awarded mainly for these two areas.

Table 5: Comparison between BAWE and BrAWE

	BAWE	BrAWE
Size	Medium-large	Medium
Quality of assignments	Merit and distinction	Passing

Source: the author

---

[10] Science without Borders (SwB) program is a mobility program created in 2011 by the government of Brazil, aiming at expanding the internationalization of Brazilian higher education.

Regarding the table above, it is worth explaining that there is a meaningful qualitative verification when it comes to the written assignments in both academic corpora. The assignments which compose the BAWE corpus are all considered merit or distinction, whereas in BrAWE they do not have a quality measurement. Therefore, BAWE is a good corpus to be used as a point of arrival.

3.4 Data analysis

In sections 3.2 and 3.3 I described the academic corpora the analysis will be based on. In this section, I will explain the steps taken to investigate the use of connectors made by Brazilian students represented in BrAWE using Sketch Engine. Due to time and space constraints, this analysis addresses only connectors of one word, such as 'moreover', rather than 'in addition'. This decision was made because the whitelist[12] can only be composed of simple expressions.

The first step was to choose a whitelist of connectors[13]. The list was then formatted into a TXT file to be read by Sketch Engine and then uploaded (appendix 1) in the "Word list" tool so that the referred connectors could be found in the corpora. The analysis was carried out separately for each corpus, i.e, BrAWE and BAWE (figure 3). The Word list in Sketch Engine searches for one-word expressions only. This way, expressions such as 'in other words' or 'for instance' had to be excluded from the search. The results are going to be further explained in chapter 4.

---

[12] A whitelist is a list of words or expressions the researcher is willing to analyze. Consequently, the outcomes show only the frequencies and the occurrences of the words in this whitelist.
[13] This list was extracted from the following website:
<http://www.maailmakeeled.ut.ee/sites/default/files/fl/linking_words_and_phrases.pdf>, a material created based on *A Grammar of Contemporary English* by Greenbaum; Leech; Svartvik (1980)

Figure 3: Word List tool in Sketch Engine

Source: Sketch Engine

In order to verify whether there are statistically significant differences in the frequencies, Log-Likelihood (LL)[14] tests were used. According to Rayson (2002) LL presents the best results to compare frequencies of words between corpora. If LL reveals 6.63 or more as an outcome, there is around 99% chance that this variation is not random but follows a certain pattern, i.e, is statistically significant.

Figure 4: Log-Likelihood calculator

Source: Log-Likelihood calculator

The results are explained in the next chapter.

---

[14] LL test can be found in the following website: <http://ucrel.lancs.ac.uk/llwizard.html>

4. RESULTS AND FINDINGS

This chapter addresses what has been found in the investigation. The table below gives a quantitative overview, containing the amount of connectors, the corpus size, the respective normalized values, and the log likelihood result. In this sense, 21,148 and 2,413 are the total amount of connectors:

Table 6: Overall figures for connectors

	BAWE	BrAWE
Total number of connectors	21,148	2,413
Corpus size – in words	6,968,089	657,859
Normalized values	3.03	3.66
Log Likelihood (LL)	- 73.92	

Source: the author

Based on these data, it is possible to affirm that there is an overuse of connectors in BrAWE comparing to the British corpus. First we have a difference in normalized values, being 3.03 occurrences per thousand words in BAWE, whereas 3.66 in BrAWE. This gives a -73.92 LL value, which is considered statistically significant.

Looking at individual connectors, the following table presents a comparison of individual occurrences between the two corpora as well as the LL result. In the first column, we have the Whitelist of connectors, the second brings the total amount in the BAWE followed by its normalized values. The fourth column presents the total amount, and the fifth the normalized values for BrAWE. The sixth and last column portrays the LL results. The highlighted lines represent the statically significant results.

As explained in the Methodology chapter, BAWE was considered Corpus 1, and BrAWE corpus 2. Thus, a "+" signal means the specific connector was more used in BAWE, whereas a "-" signal means the connector was more used in BrAWE. Additionally, when LL figures were higher than 6.63 (RAYSON, 2002), it means the difference was statistically significant, i.e., not a random choice, but a pattern of use.

Table 7: Comparison of individual connectors between BAWE and BrAWE

Whitelist	BAWE	Normalized value	BrAWE	Normalized value	LL
However	7,857	1.127	658	1.000	+ 9.02
Therefore	2,700	0.387	359	0.545	- 34.05
Thus	1,673	0.240	130	0.197	+ 4.84

Furthermore	1,214	0.174	161	0.244	- 15.05
Moreover	989	0.141	179	0.272	- 55.13
Hence	858	0.123	30	0.045	+ 39.59
Finally	754	0.108	159	0.241	- 70.89
Then	673	0.096	162	0.246	- 93.71
Firstly	557	0.079	71	0.107	- 5.25
Similarly	512	0.073	20	0.030	+ 19.92
Nevertheless	458	0.065	88	0.133	- 31.66
Consequently	418	0.059	48	0.072	- 1.57
Secondly	414	0.059	38	0.057	+ 0.03
Instead	371	0.053	13	0.019	+ 17.07
Additionally	312	0.044	54	0.082	- 14.65
Besides	204	0.029	127	0.193	- 218.40
Nonetheless	155	0.022	17	0.025	- 0.33
Conversely	152	0.021	9	0.013	+ 2.13
Initially	130	0.018	17	0.025	- 1.47
Lastly	128	0.018	11	0.016	+ 0.09
Likewise	116	0.016	12	0.018	- 0.09
Meanwhile	115	0.016	12	0.018	- 0.11
Thirdly	99	0.014	9	0.013	+ 0.01
Accordingly	81	0.011	6	0.009	+ 0.35
Subsequently	70	0.010	13	0.019	- 4.29
Eventually	62	0.008	0	0	+ 11.19
Otherwise	48	0.006	10	0.015	- 4.34
Simultaneously	13	0.001	0	0	+ 2.35
Thereafter	10	0.001	0	0	+ 1.80
Meantime	5	0.0007	0	0	+ 0.90
Afterward(s)	0	0	0	0	+ 0.00

From the 31 connectors of the whitelist, 13 showed up statistically different. Five of them are more frequent in BAWE: 'however', 'hence', 'similarly', 'instead', and 'eventually'. On the other hand, eight are overused by Brazilian students in the BrAWE when compared to written texts assigned merit or distinction grades which compose the BAWE corpus: 'therefore', 'furthermore', 'moreover', 'finally', 'then', 'nevertheless', 'additionally', and 'besides'.

Considering types of connectors, there does not always seem to be a pattern of overuse by Brazilian students. For example, whereas 'however', a contrast connector, according to Biber et al (1999) has a higher frequency in BAWE, another contrast

connector, 'nevertheless', is a lot more frequent in BrAWE. Following the same fashion, 'hence', from the result/inference category is more frequent in BAWE, while 'therefore' and 'then' are overused in BrAWE. On the other hand, the Enumeration and Addition category is indeed overused in BrAWE, with 'furthermore', 'moreover', 'additionally', and, above all, 'besides', with a -218.4, the highest LL value in the data. The table below systematizes the statistically significant results regarding categories of connectors in both corpora:

Table 8: Categories of connectors in both corpora

CATEGORY	BAWE	BrAWE
Contrast	However	Nevertheless
Result/Inference	Hence	Therefore, then
Enumeration/Addition	---	Furthermore, moreover, additionally, besides

Source: the author

Furthermore, four connectors from the whitelist used in this investigation only occur in the BAWE corpus. It is the case of 'eventually' (already mentioned as a statistically significant underused connector), 'simultaneously', 'thereafter' and 'meantime', whose frequencies in the British corpus are 62, 13, 10 and 5 respectively. 'Afterward(s)' is the only connector from the whitelist that does not occur in both corpora analyzed.

The outcomes of this study follow similar patterns to the results presented in the literature review (GRANGER; TYSON, 1996; MILTON, 2001; NARITA; SATO; SUGIURA, 2004). French, Chinese and Japanese learners of English overuse the connectors 'moreover' and underuse 'however'. Taking the entire range of connectors into account, there is a high ratio of overuse by learners of various mother-tongue backgrounds. Hong Kong students, as well as Cantonese and Norwegian mother-tongue learners use far more connectors then English-speaking counterparts. (MILTON; TSANG, 1993; FIELD; YIP, 1992; EVENSEN; RYGH, 1988).

Granger and Tyson (1996, p. 19) claim that learners transfer language patterns from one language into another: "[i]f transfer plays a role in foreign language production, and it is nowadays almost universally recognized that it does, then this leads us to suggest that we will find a general overuse of connectors in the French students'

writing.". In addition, texts in French usually have a higher frequency of connectors than texts in English. Thus, texts written in French and translated into English will have omission of connectors in order to avoid stilted texts. (HERVEY; HIGGINS, 1992, apud GRANGER; TYSON, 1996). Considering both French and Portuguese are Latin languages, and French students overuse connectors as a whole, then this might explain the overuse of connectors in BrAWE as well. 'Besides' and 'then' are the connectors with the highest LL values in BrAWE - 218.40 and - 93.71 respectively. A possible explanation for that might have to do with an influence of the Portuguese language, since the Portuguese equivalents for 'besides' (*além disso*) and 'then' (*então*) might be commonly used by Brazilian students. However, a more detailed investigation should be undertaken in order to validate this hypothesis.

5. CONCLUSIONS

The objective of this study was to conduct a corpus-based quantitative analysis on how Brazilian students use connectors in their written assignments in academic contexts. Therefore, two research questions guided this investigation:

a. How do Brazilian students in the UK use connectors in their written assignments in English?

b. Do these uses differ from students represented in the British Academic Written English (BAWE) corpus?

The research findings suggest that there is an overall overuse of connectors in the corpus containing assignments from Brazilian students, i.e., BrAWE when compared to BAWE. From the total amount of words in BAWE (6,968,089), there are 21,148 occurrences of connectors, a normalized value of 3.03. In turn, the number of occurrences in BrAWE is 2,413, whose normalized value is 3.66. The LL difference for the group is -73.92. On the other hand, not all connectors are statistically overused, as is the case of 'however', 'hence', 'similarly', 'instead', and 'eventually', which are underused in BrAWE when compared to BAWE.

Regarding the English classroom and the teaching of connectors, some pedagogical implications need to be presented. Catalán and Alba (2014) state how difficult learning connectors is. Their study upholds the importance of focusing on these linguistic aspects in an English classroom environment. However, "we should not forget that connectors represent one small aspect of cohesion. Although we need to teach their correct use, we should not do so to the detriment of other forms of cohesion." (GRANGER. TYSON, 1996, p. 26). Nevertheless, considering the motivation for conducting this analysis derives from my experience as an English teacher at the LwB program, the results can be applied in the EAP lessons, since academic writing is a growing demand in this context. Any teacher can use corpus linguistics tools with their students, because they provide evidences of language production in texts that were not created for pedagogical purposes (SARMENTO, 2010).

As suggestion for follow up studies, connectors which contain more than one word should be analyzed. Due to time constraints, this investigation only analyzed connectors containing one word (see appendix 1) as the Word list tool enables searches

for one word only. Also, and most importantly, further studies about the behavior of the connectors in the texts should be carried out, since only frequency was analyzed here. Position in the sentence and collocations are possible objects of study in order to verify how connectors are used by Brazilians in comparison to students represented in the BAWE corpus.

REFERENCES

ALSOP, S.; NESI, H. Issues in the development of the British Academic Written English (BAWE) corpus. *Corpora*, v. 4, n. 1, p. 71-83, 2009.

BARROS, D. *Teoria Semiótica do Texto*. São Paulo: Ática, 1990.

BAUMAN, J. F; GRAVES, M. F. What is academic vocabulary ?. Journal of Adolescent and Adult Literacy, v. 54, n. 1, p. 4 – 12, 2010.

BEAUGRANDE, R. A. *New Foundations for a Science of Text and Discourse*. Greenwich, CT: Ablex, 1997.

BERMAN, R.; CHENG, L. English academic language skills: Perceived difficulties by undergraduate and graduate students, and their academic achievement. *Canadian Journal of Applied Linguistics/Revue canadienne de linguistique appliquée*, v. 4, n. 1, p. 25-40, 2010.

BIBER, D. et al. *Longman Grammar of Spoken and Written English*. Longman, 1999.

CATALÁN, R. M. J.; ALBA, J. O. Diagnosis of EFL Learners' Difficulties in the Use of Connectors in a writing task: Strategies for Remedial Work. *Didáctica, Lengua y Literatura*, v. 26, p. 197-216, 2014.

CLARK, M.; ISHIDA, S. Vocabulary knowledge differences between placed and promoted EAP students. *Journal of English for academic purposes*, v. 4, p. 225–238, 2005.

COBB, T.; HORS, M. Learner Corpora and Lexis. In. S. Granger; G. Gilquin; F. Meunier (Eds.), *The Cambridge Handbook of Learner Corpus Research*. Cambridge: Cambridge University Press, 2015.

COXHEAD, A. The Academic Word List 10 years on: research and teaching implications. *TESOL Quarterly,* v. 4, n. 2, p. 355 – 362, 2011.

EASTWOOD, J. *Oxford Guide to English Grammar*. Oxford University Press, 1994.

EVENSEN, L. S.; RYGH, I. L. Connecting L1and FL in discourse-level performance analysis. *Papers and Studies in Contrastive Linguistics*, v. 22, p. 133 – 178, 1988.

FIELD, Y.; YIP, L. M. O. A Comparison of Internal Conjunctive Cohesion in the English Essay Writing of Cantonese Speakers and Native Speakers of English. *RELC Journal,* v. 23, n. 1, p. 15 – 28, 1992.

FLOWERDEW, L. The argument for using English specialized corpora to understand academic and professional language. In U. Connor; T. Upton (Eds.), *Discourse in the professions*: *Perspectives from corpus linguistics,* p. 11-36. Amsterdam: John Benjamins Publishing Company, 2004.

GOLDMAN, S. R.; MURRAY, J. D. Knowledge of Connectors as Cohesion Devices in Text: A Comparative Study of Native-English and English-as-a-Second-Language Speakers. *Journal of Educational Psychology*, v. 84, n. 4, p. 504-519, 1992.

GRANGER, S.; TYSON, S. Connector usage in the English essay writing of native and non-native EFL speakers of English. *World, Englishes,* v. 15, n. 1, p. 17–27, 1996.

HUTCHINSON, T.; WATERS, A. English for Specific Purposes: A learning-centered approach. Cambridge University Press, 1987.

HYLAND, K. General and specific EAP. In: Hyland, Ken; Shaw, Philip. The Routledge Handbook of English for Academic Purposes, London, Routledge, 2016.

HYLAND, K.; HAMP-LYONS, L. EAP: Issues and Directions. *Journal of English for Academic Purposes*, v.1, n.1, p. 1-12. 2002.

LAMBERTS, D. V. H. *O livro didático de língua inglesa em uso: análise de pesquisas e observações de aula no programa idiomas sem fronteiras-inglês.* Dissertação de Mestrado. Universidade Federal do Rio Grande do Sul, 2015.

MCENERY, T.; HARDIE, A. *Corpus linguistics: Method, theory and practice.* Cambridge: Cambridge University Press, 2011.

MILTON, J. *Elements of a written interlanguage: A computational and corpus-based study of institutional influences on the acquisition of English by Hong Kong Chinese students.* Language Centre: The Hong Kong University of Science and Technology, 2001.

MILTON, J.; TSANG, E. S. C. A Corpus-based study of logical connectors in EFL students' writing: directions for future research. In *Studies in Lexis*. Edited by R. Pemberton and E.S.C. Tsang. Hong Kong: The Hong Kong University of Science and Technology. p. 215-246, 1993.

NARITA, M.; SATO, C.; SUGIURA, M. Connector Usage in the English Essay Writing of Japanese EFL Learners. *Language, Resources and Evaluation*, p. 1171 – 1174, 2004.

NATION, P. *Learning vocabulary in another language.* Cambridge: Cambridge University Press, 2013.

PLANCIC, B.; NINCEVIC, S. Transition Words in Academic Writing. *Transactions on Maritime Science*, v. 1, p. 61-67, 2014.

RAYSON, P. *Matrix: A statistical method and software tool for linguistic analysis through corpus comparison*. PhD Thesis, Lancaster University, 2002.

SARDINHA, A. Linguística de corpus: histórico e problemática. *D.E.L.T.A.* v. 16, n. 2, p. 323–367, 2000.

SARMENTO, S. Linguística de corpus: histórico, metodologia, campos de aplicação. *Revista Trama*, v. 6, n. 2, p. 87-107, 2010.

SARMENTO, S. et al. Do Inglês sem Fronteiras ao Idiomas sem Fronteiras: A construção de uma política linguística para a internacionalização. Belo Horizonte: Editora UFMG, 2016.

SILVA, L. G. *Academic vocabulary: a corpus linguistics study on how Brazilian students write academic English*. Master's Thesis. University of Warwick, 2016.

TAPPER, M. Connectives in advanced Swedish EFL learners' written English – preliminary results. *The Department of English: Working Papers in English Linguistics*, v. 5, 116-144, 2005.

VIAL, A. P. S. *O uso do livro didático no programa inglês sem fronteiras da Universidade Federal do Rio Grande do Sul*. Trabalho de Conclusão de Curso. Universidade Federal do Rio Grande do Sul, 2014.

APPENDICES

Appendix 1 – Whitelist

Afterward(s)
Initially
Meanwhile
Meantime
Simultaneously
Eventually
Subsequently
Then
Thereafter
Firstly
Secondly
Thirdly
Finally
Lastly
Conversely
However
Instead
Nevertheless
Nonetheless
Otherwise
Likewise
Similarly
Additionally
Besides
Furthermore
Moreover
Accordingly
Consequently
Hence
Therefore
Thus

Appendix 2 – Types of connectors in BAWE (in order of frequency)

However
Therefore
Thus
Furthermore
Moreover
Hence
Finally
Then
Firstly
Similarly
Nevertheless
Consequently
Secondly
Instead
Additionally
Besides
Nonetheless
Conversely
Initially
Lastly
Likewise
Meanwhile
Thirdly
Accordingly
Subsequently
Eventually
Otherwise
Simultaneously
Thereafter
Meantime

Appendix 3 – Types of connectors in BrAWE (in order of frequency)

However
Therefore
Moreover
Then
Furthermore
Finally
Thus
Besides
Nevertheless
Firstly
Additionally
Consequently
Secondly
Hence
Similarly
Nonetheless
Initially
Subsequently
Instead
Meanwhile
Likewise
Lastly
Otherwise
Thirdly
Conversely
Accordingly