

## **A corpus-based study of connectors in student academic writing**

**Marine Laísa Matte, Simone Sarmento**

### **Abstract:**

The aim of this paper is to present the results of an investigation which dealt with 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 use, and help in the creation of pedagogical material for English language classrooms.

**Key words:** Corpus Linguistics; English for Academic Purposes; Connectors.

1

### **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 on the vocabulary used in these texts, coupled with other aspects, naturally. 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.

There are some words and expressions, such as connectors<sup>1</sup>, that are essential to all users wishing to master writing skills in an additional language (Plancic; Nincevic, 2014). Even though connectors are considered to be grammar words that do not fit into the classification of general academic vocabulary<sup>2</sup>, 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 a corpus composed of written assignments in English produced by Brazilian students whose first language is Portuguese.

The motivation to conduct a study on written English derives from our experience as instructors in English for Academic Purposes courses at a university in the south of Brazil. The teaching of Academic English in Brazilian universities is a relatively new educational phenomenon, which arose with the onset of the government sponsored mobility program called Science without Borders (SwB)<sup>3</sup>. With the language issues<sup>4</sup> which stemmed from SWB, the Languages without Borders (LwB) Program was created with the aim of providing English classes to university undergraduate and graduate students, professors and administrative staff. Most of these courses have a very specific objective, and have as a final aim to enhance the quantity and the quality of Brazilian academic publications in English.

The LwB assists students from all levels of proficiency. When marking assignments, students revealed a lack of familiarity with the most common words that are used to connect ideas. Connectors 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). This way, the main aim of this study is to verify

---

<sup>1</sup> 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)

<sup>2</sup> 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)

<sup>3</sup> Science without Borders – **SwB** was the international student mobility program devised, implemented and funded in 2011 by the Brazilian federal government. Its main objective was to promote the internationalization of Brazilian science and technology by providing complementary qualification of Brazilian higher education students in universities of excellence abroad. Its target was to send, over the period of four years, 101,000 undergraduate and graduate students to attend sciences, technology, engineering and mathematics-related courses in universities, in 29 other countries.

<sup>4</sup> Lack of proficiency in English was reported as the main obstacle for reaching the 101,000 target.

the use of connectors in a corpus composed of assignments written by Brazilians studying in British universities (BrAWE)<sup>5</sup>. The results will be compared to a corpus composed by texts written by a group of students with high marks in British universities (BAWE). The questions guiding the research are:

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

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

## 2. Literature review

### 2.1 English for Academic Purposes (EAP)

Emerging from the field of English for Specific Purposes (ESP) (see figure 1), English for Academic Purposes (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). Thus, the teaching of EAP is for those students whose needs are related to uses in academic contexts. That is, based on the demands of a certain area, students can look for academic English classes, for instance, to learn how to write a specific genre (Gardner; Nesi, 2013), to learn more academic vocabulary or also to be exposed to technical vocabulary related to their own area of expertise.

The figure below depicts different branches of English language teaching, including EAP from English for General Purposes (EGP):

---

<sup>5</sup> BAWE and BrAWE will be described later on this text.

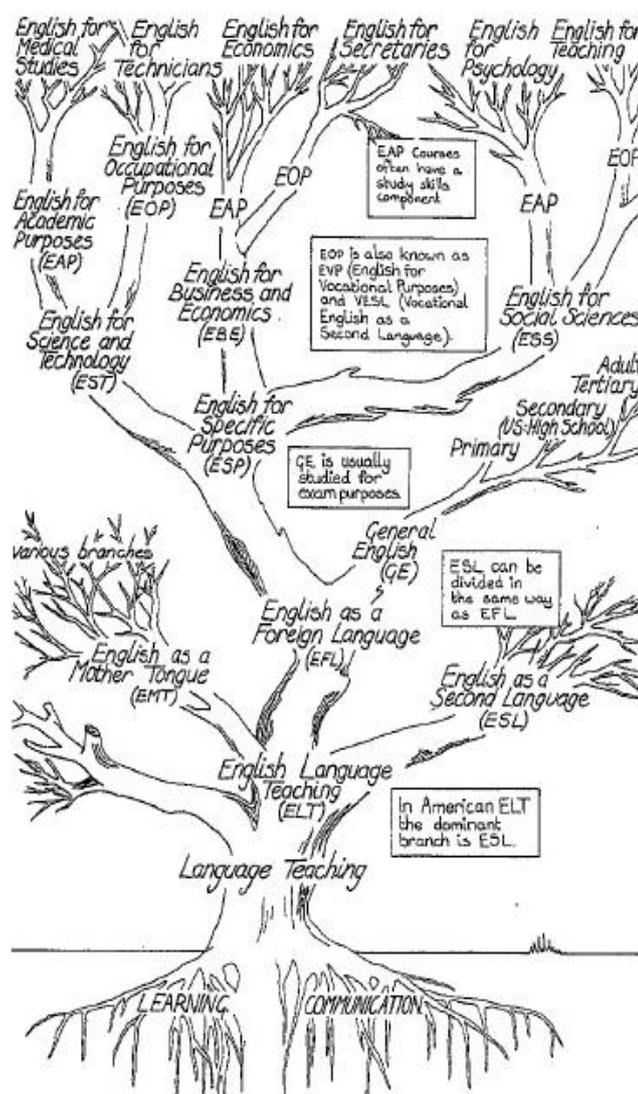


Figure nº 1 Hutschinson and Waters' ELT Tree (1987)

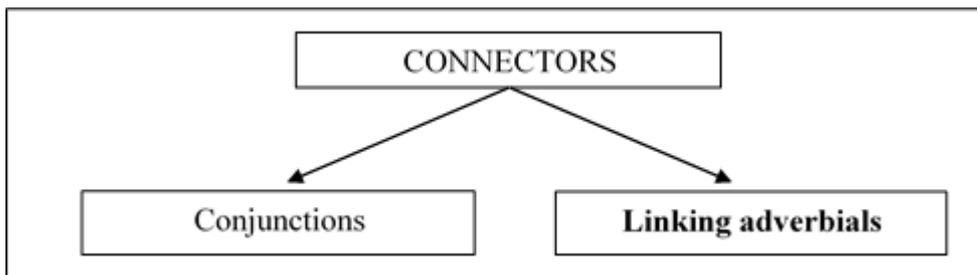
From the picture above, it is possible to observe that learning and communication compose the roots of the Language Teaching tree. In turn, different branches emerge from the English Language Teaching tree. EAP derives from ESP, which, derives from English as a Foreign Language (EFL), diverging from 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, the author brings up another subdivision in the EAP field, since there is the general EAP, or English for

General Academic Purposes (EGAP), and the specific EAP, what he calls English for Specific Academic Purposes (ESAP). This segmentation becomes productive when different teaching contexts are taken into account. Some courses will only deal with students from the same field, and then, ESAP will be the approach. Conversely, when you have students from a myriad of areas, only EGAP will be possible. The focus here is only on EGAP, which is the most common situation regarding English classes in Brazilian universities. 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. Hence, this study does not take into consideration specificities of different areas of expertise and will treat both BAWE and BrAWE as one group of texts only.

## 2.2 Connectors

As stated in the Longman Grammar of Spoken and Written English (LSWE), *connectors* are not a grammatical group of words. Instead, some specific 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 nº 2 - Connectors as an umbrella term**

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)

6

<b>Category</b>	<b>Examples</b>
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

### **Table n° 1 – Biber et al's linking adverbials categories (1999)**

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 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)<sup>6</sup> 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 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

---

<sup>6</sup> Translation made by the authors. Original quote: “[...]o texto só existe quando concebido na dualidade que o define — objeto de significação e objeto de comunicação [...]” (BARROS, 1990, p. 12).

relation 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)

8

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

Several studies have been conducted to analyze the English language learners uses of connectors (Granger; Tyson, 1996; Milton, 2001; Narita; Sato; Sugiura, 2004; Tapper, 2005), but, none has been found in terms of how Brazilians use connectors in English. 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: 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, 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 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: ‘however’, ‘although’ and ‘for example’.

In the same line, based on a quantitative analysis of connectors with advanced Japanese students’, Narita, Sato and Sugiura (2004) compared their essays in English with the essays written by native English speakers. In their findings additional connectors tend to be overused, such as ‘moreover’ and ‘in addition’. Regarding the underused connectors, the authors mention ‘however’ and ‘then’. This study goes further and tries to explain why contrastive connectors are underused 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)

Tapper (2005) compared Swedish advanced learners of 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 clarifying connectors such as ‘that is’, ‘in other words’, ‘for example’, which, according to Biber et al (1999), are called 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:

<b>Study</b>	<b>Overuse</b>	<b>Underuse</b>
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	-----

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

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, learners underuse connectors such as ‘however’ and ‘therefore’. Narita, Sato and Sugiura’s study (2004) took the relation between language use and proficiency into account. However, in the present study, proficiency level is not considered, as Brazilian students in the BrAWE corpus are all competent users of English, as will be seen in the section below.

### **3. Methodology**

As previously mentioned, this study uses a corpus linguistics approach to investigate how Brazilian students use connectors in their written assignments across British universities. In this section, we will present our views on how corpus linguistics was used as the methodological approach and describe the corpora used for this study. In addition, we will explain the steps taken to analyze the corpus.

#### **Corpus linguistics**

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.” Thus, corpus linguistics is a resource whose compilation of texts in digital format allows the researcher to explore uses of the language in real contexts. It

focuses on language exploitation through empirical evidences, extracted from a computer. In addition,, 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)<sup>7</sup>, 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”.

### **British Academic Written English (BAWE) and Brazilian Academic Written English (BrAWE)**

As already mentioned, this investigation uses two academic English corpora to analyze the uses of connectors: BAWE and BrAWE.

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 a later few at Coventry University. The corpus is composed of written assignments from students regardless of their nationalities. These assignments were all written in English and submitted electronically. Moreover, all the assignments were judged as standard productions in their areas (Alsop; Nesi, 2009); i.e., they were all merit and distinction assignments.

BAWE can be considered a medium-large-size corpus, as it has 6,968,089 tokens<sup>8</sup>. For Sardinha (2000), a corpus with 1 million to 10 million tokens is medium-large.

Alsop and Nesi (2009, 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

---

<sup>7</sup> Translated by the authors. 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)

<sup>8</sup> 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.

Sciences (PS). Although it is allegedly a balanced corpus, it contains more texts in AH and SS when compared to PS and LS. As reported by 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.” Hence, this corpus has more detailed information when compared to the other corpus used in this investigation

BrAWE (Silva, 2016) is a medium-size corpus (Sardinha, 2000), as it contains 657,859 tokens. 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.

BrAWE has also written assignments from the same four areas: Social Sciences (SS), Arts and Humanities (AH), Physical Sciences (PS), and Life Sciences (LS). Apart from the difference in size, there are other aspects that distinguish both corpora. , The first is that, “[...] 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 (BrAWE) here are significantly smaller than the other two [...]” (Silva, 2016, p. 42). This difference comes from the fact that BrAWE was compiled during the SwB which only catered for students from STEM<sup>9</sup> and Health Sciences students. Also, whereas BAWE was composed of highly graded assignments only, BrAWE accepted assignments graded pass as well. Therefore, BAWE is a good corpus to be used as a point of arrival.

	<b>BAWE</b>	<b>BrAWE</b>
Size	6,968,089 Medium-large	657,859 Medium
Quality of assignments	Merit and distinction	Passing

**Table n° 3: Comparison between BAWE and BrAWE**

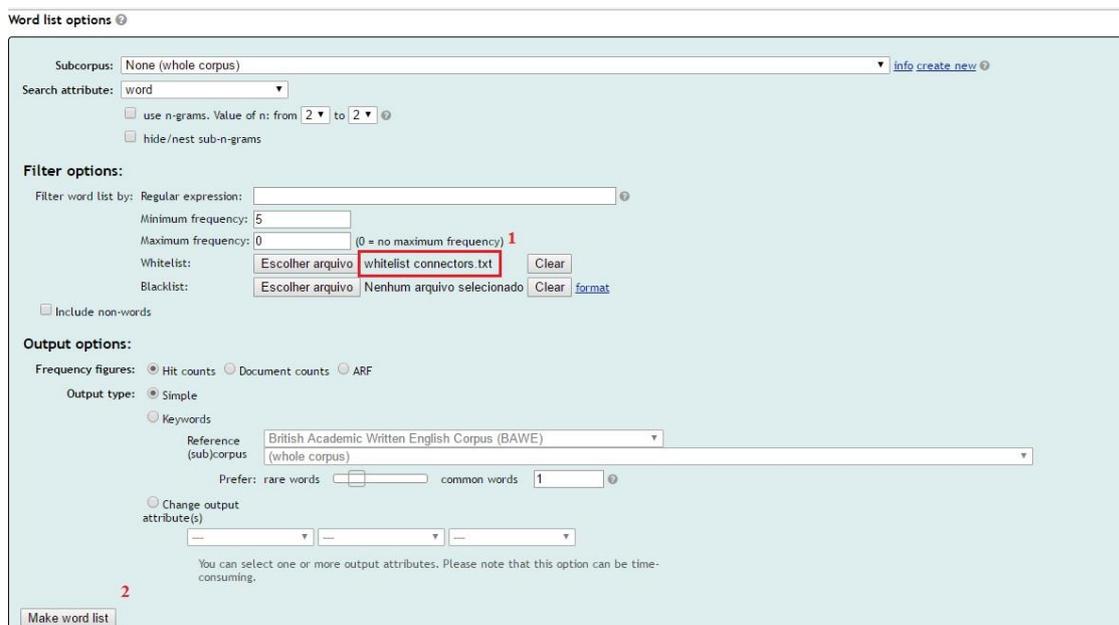
### Methodological Procedures

The analysis was carried out using only one word connectors, such as ‘moreover’, rather than ‘in addition’. This decision was taken because the whitelist<sup>10</sup> can only be composed of simple expressions.

<sup>9</sup> Science, technology, engineering, and mathematics.

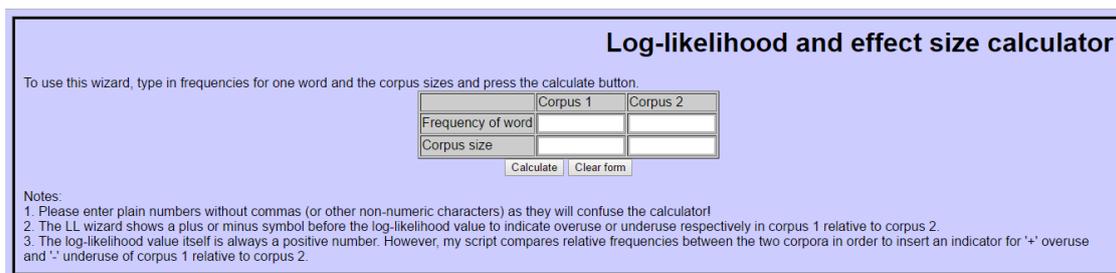
<sup>10</sup> A whitelist is a list of words or expressions the researcher is willing to analyze. Consequently, the outcomes show only the frequencies and the occurrences of the words in this whitelist. This list was extracted from the following website: <[http://www.maaillmakeeled.ut.ee/sites/default/files/fl/linking\\_words\\_and\\_phrases.pdf](http://www.maaillmakeeled.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)

The first step was to choose a whitelist of connectors. 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). As the Word list in Sketch Engine searches for one-word expressions only, expressions such as ‘in other words’ or ‘for instance’ had to be excluded from the search.



**Figure nº 3: Word List tool in Sketch Engine**

In order to verify whether there are statistically significant differences in the frequencies, the Log-Likelihood (LL)<sup>11</sup> test was 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.



<sup>11</sup> LL test can be found in the following website: <<http://ucrel.lancs.ac.uk/llwizard.html>>

**Figure n° 4: Log-Likelihood calculator****4. Results and findings**

The table below gives a quantitative overview, containing the corpus size, the amount of connectors, the respective normalized values, and the log likelihood result. In this sense, 21,148 and 2,413 are the total amount of connectors:

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

**Table n° 4: Overall figures for connectors**

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, Table 5 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 show the statistically significant results.

As explained before, BAWE was considered Corpus 1, and BrAWE corpus 2. Thus, a “+” signal means the specific connector was more frequent in BAWE, whereas a “-” signal means the connector was more frequent 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.

<b>Whitelist</b>	<b>BAWE</b>	<b>Normalized value</b>	<b>BrAWE</b>	<b>Normalized value</b>	<b>LL</b>
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

**Table nº 5: Comparison of individual connectors between BAWE and BrAWE**

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 the Brazilian students. For example, whereas ‘however’, a contrast connector (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:

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

**Table n° 6: Categories of connectors in both corpora**

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 as 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 than 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 can be. 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 our experience English instructors in academic contexts, the results can be applied in EAP lessons, since academic writing is a growing demand in this context.

As suggestion for follow up studies, connectors which contain more than one word should be analyzed, as we only analyzed connectors containing one word due to the Word list limitation which enables searches for one word. 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:

1. 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.
2. BARROS, D. *Teoria Semiótica do Texto*. São Paulo: Ática, 1990.
3. BAUMAN, J. F; GRAVES, M. F. What is academic vocabulary ?. *Journal of Adolescent and Adult Literacy*, v. 54, n. 1, p. 4 – 12, 2010.
4. BEAUGRANDE, R. A. *New Foundations for a Science of Text and Discourse*. Greenwich, CT: Ablex, 1997.
5. BIBER, D. et al. *Longman Grammar of Spoken and Written English*. Longman, 1999.
6. 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.

7. 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.
8. COXHEAD, A. The Academic Word List 10 years on: research and teaching implications. *TESOL Quarterly*, v. 4, n. 2, p. 355 – 362, 2011.
9. EASTWOOD, J. *Oxford Guide to English Grammar*. Oxford University Press, 1994.
10. EVENSEN, L. S.; RYGH, I. L. Connecting L1 and FL in discourse-level performance analysis. *Papers and Studies in Contrastive Linguistics*, v. 22, p. 133 – 178, 1988.
11. 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.
12. 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.
13. 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.
14. 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.
15. HYLAND, K. General and specific EAP. In: Hyland, Ken; Shaw, Philip. *The Routledge Handbook of English for Academic Purposes*, London, Routledge, 2016.
16. HYLAND, K.; HAMP-LYONS, L. EAP: Issues and Directions. *Journal of English for Academic Purposes*, v.1, n.1, p. 1-12. 2002.
17. MCENERY, T.; HARDIE, A. *Corpus linguistics: Method, theory and practice*. Cambridge: Cambridge University Press, 2011.
18. 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.
19. 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.
  20. 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.
  21. NATION, P. *Learning vocabulary in another language*. Cambridge: Cambridge University Press, 2013.
  22. PLANCIC, B.; NINCEVIC, S. Transition Words in Academic Writing. *Transactions on Maritime Science*, v. 1, p. 61-67, 2014.
  23. RAYSON, P. *Matrix: A statistical method and software tool for linguistic analysis through corpus comparison*. PhD Thesis, Lancaster University, 2002.
  24. SARDINHA, A. Linguística de corpus: histórico e problemática. *D.E.L.T.A.* v. 16, n. 2, p. 323–367, 2000.
  25. SARMENTO, S. Linguística de corpus: histórico, metodologia, campos de aplicação. *Revista Trama*, v. 6, n. 2, p. 87-107, 2010.
  26. SILVA, L. G. *Academic vocabulary: a corpus linguistics study on how Brazilian students write academic English*. Master's Thesis. University of Warwick, 2016.
  27. 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.

### **Marine Laísa Matte**

Master student in Applied Linguistics  
Universidade Federal do Rio Grande do Sul, Brazil  
[marine.laisa@gmail.com](mailto:marine.laisa@gmail.com)

#### **Area of research**

- English for Academic Purposes
- Corpus Linguistics

### **PhD Simone Sarmento**

Professor from Applied Linguistics Graduate Department

A corpus-based study of connectors in student academic writing  
*Marine Laísa Matte, Simone Sarmento*

Universidade Federal do Rio Grande do Sul, Brazil

[simone.sarmiento@ufrgs.br](mailto:simone.sarmiento@ufrgs.br)

**Area of research**

- English for Academic Purposes
- Corpus Linguistics
- Internationalization of Higher Education
- Language Teacher Education