

## Configuring Collocations and Prepositions in Essay Writing through a Corpus-based Strategy

### Estructuración de Colocaciones y Preposiciones en Escritura de Ensayos Usando un Corpus Lingüístico

Leonardo Alba-López<sup>1</sup>

*Indiana University, Bloomington, Indiana, United States*

#### Abstract

The correct use of collocations and prepositions plays an essential role in writing. Previous research has demonstrated that students, who master these lexical elements, have a better language awareness, improve their reading and listening comprehension, and are more creative when making a composition. However, little specific work has been done in this regard in Colombia, especially with adolescents. This action research study explores the affordances of a corpus in English as a foreign language writing of fourteen students at the high school level. Surveys, teacher's journals, and classroom artifacts were used to collect data. Findings show that the corpus contributed positively to L2 writing in two ways. The participants could self-correct their mistakes associated with the use of collocations and prepositions, and the tool amplified participants' autonomy, decision-making, and data analysis skills. Although a corpus represents an alternative for error correction processes, it requires a well-structured instructional design process to deal with difficulties related to using commands derived from a corpus-based strategy.

*Keywords:* academic writing, collocations, corpus linguistics, prepositions

---

<sup>1</sup> He is a first-year doctoral student in the Instructional Systems Technology program at Indiana University. He earned his undergraduate degree in foreign languages at Universidad Surcolombiana, Colombia. He has two Master's degrees, one in English Language Teaching and Autonomous Learning Environments at Universidad de La Sabana (Chia, Colombia); and another in Education at Carthage College (Kenosha, Wisconsin). His interests are technology integration into educational settings, digital literacy, and education research. He has recently worked as an adjunct teacher for a modern languages program, teaching courses, such as material design and research methods. Currently, he is an associate instructor at Indiana University. [laalbalopez@iu.edu](mailto:laalbalopez@iu.edu)

ORCID: <https://orcid.org/0000-0003-1444-7156>

Received: June 6<sup>th</sup>, 2023. Accepted: June 13<sup>th</sup>, 2024

This article is licensed under a Creative Commons Attribution-Non-Commercial-No-Derivatives 4.0 International License. License Deed can be consulted at <https://creativecommons.org/licenses/by-nc-nd/4.0>.

## Resumen

El uso correcto de colocaciones y preposiciones tiene un papel primordial en la escritura. Investigaciones previas han demostrado que los estudiantes que dominan estos elementos léxicos tienen mejor conciencia del idioma, mejoran su comprensión de lectura y escucha, y son más recursivos al escribir. Sin embargo, pocos estudios relacionados con el uso de colocaciones y preposiciones se han hecho en Colombia. La presente investigación-acción tiene como objetivo explorar el uso de un corpus lingüístico en la escritura en inglés como lengua extranjera de estudiantes de secundaria. Los métodos de recolección de datos incluyen diarios del docente, encuestas y escritos de los estudiantes. Los resultados de esta investigación contribuyen en la corrección de errores de escritura en inglés de dos maneras. Primero, los estudiantes autocorrigieron sus errores relacionados con colocaciones y preposiciones utilizando un corpus lingüístico. Asimismo, la implementación del corpus lingüístico en procesos de corrección de escritos potenció la autonomía de los estudiantes y fortaleció sus habilidades de toma de decisiones y análisis de información. Aunque un corpus lingüístico representa una alternativa para procesos de corrección de errores, su uso requiere un diseño instruccional estructurado para contrarrestar las dificultades relacionadas con el uso de los comandos que se utilizan en los corpus lingüísticos.

*Palabras clave:* colocaciones, corpus lingüístico, preposiciones, escritura académica

## Introduction

Several factors interplay in the process of learning English as a Foreign Language (EFL). Learners must learn syntax, vocabulary, and phonetics to master and use a foreign language competently in real-life settings. They also go through an acculturation process and develop a new way of thinking, feeling, and acting (Brown, 2007; Hoff & Core, 2013). Mastering the four linguistic skills (listening, reading, writing, and speaking) requires a considerable amount of time and implies commitment and ongoing practice. Even though a simultaneous development of the four linguistic skills is desirable, reading and writing are the linguistic abilities that demand a longer process as they both are part of the advanced stages of language development (Hoff & Core 2013; Kennison, 2014; Stern, 1970).

Learning to write effectively is an ability that learners typically find complicated and tumultuous, which might be attributed to different factors. While Tribble (1996) pointed out that there is a natural refusal to write in our mother tongue or any other language, Benko (2012) argued that inappropriate curriculum and teachers, who are insufficiently trained for teaching students writing, affect writing development. Moreover, a lack of practice in modeling, telling, retelling, or transforming pieces of information also prevents students from acquiring writing competence (Benko, 2012; Graham & Perin, 2007; Myles, 2002).

When teaching writing, it is key to consider its types (e.g., creative, social, personal, etc.) as each type has its conventions. However, educators from schools to higher education levels typically put aside academic writing development due to its complexity. Academic writing requires the writer to be informed while possessing higher-order thinking skills

and a sustained commitment to the writing task (Listyani, 2018; Myles, 2002; Oshima & Hogue, 2007). It also requires integration of audience, purpose, organization, style, flow, and presentation (Swales & Feak, 2004). The latter, which is related to the identification of different structural mistakes through proofreading, is one of the concerns of this study.

While conventions of academic writing may be one issue, information and communication technologies (ICTs, hereafter) have brought several changes as they have had a critical influence on the ways we create written content, communicate, and engage with other writers and readers (Hyland, 2002). Indeed, with the ascendance of ICTs as content media, technology has become a multi-purpose tool in learners' personal and professional lives (Dudeny & Hockly, 2007). Thus, new technologies influence our lives unquestionably.

A corpus which is a collection of spoken or written examples of a language in which one can find word frequencies, concordance lines of words, and excerpts of texts (O'Keeffe et al., 2007) is a technological tool that has been used in multiple fields, including education. Numerous research studies have shown the benefits of using corpora in educational settings. These include the making of English language dictionaries and textbook dialogues, and the improvement of learners' language awareness (Biber et al., 1998; Granger, 1998; Johns, 1991; Kettemann, 1995). Nevertheless, many scholars have questioned corpora usage in those contexts as they wonder whether the sample of texts found in a corpus, which are withdrawn from their original context, can be considered suitable for educational purposes.

Little specific work has been done on implementing a corpus, like the Corpus of Contemporary American English (COCA), in educational settings in Colombia. Thus, this study aims to determine the effectiveness of a corpus-based strategy combined with the writing process approach to improve the configuration of collocations and prepositions in essay writing in an EFL context amongst adolescents.

Thus, to confirm the impact, whether positive or negative, of implementing a corpus-based strategy for error correction, the following specific objectives have been established: (a) to determine the advantages and disadvantages of using this type of tool from the students and teacher's perspective, and (b) to determine how learners become autonomous when correcting their writing mistakes throughout the use of a corpus.

### *Qualitative Question*

What does incorporating the Corpus of Contemporary American English (COCA) in the writing process tell us about the configuration of collocations and prepositions in essay writing in the EFL context?

## *Hypotheses*

This study investigated the following hypotheses:

### *Hypothesis one:*

- $H_0$ : Implementing a corpus-based strategy will increase the mistakes between the first and second versions of writing task 1.
- $H_1$ : Implementing a corpus-based strategy will decrease the mistakes between the first and second versions of writing task 1.

### *Hypothesis two:*

- $H_0$ : The average of collocation and preposition mistakes of the Pre-test is equal to the Post-test ( $\mu_{Pre} = \mu_{Post}$ ).
- $H_1$ : The average of collocation and preposition mistakes of the Pre-test is not equal to the Post-test ( $\mu_{Pre} \neq \mu_{Post}$ ).

## *Mixed Methods Question*

Do the qualitative and quantitative data converge to support the conclusion that a corpus-based strategy combined with the writing process approach improves the configuration of collocations and prepositions in essay writing in EFL scenarios and learners' autonomy when correcting their compositions?

## Definition of Terms

### *Collocation*

It is how words combine in a language to produce natural-sounding speech and writing. For example, in English, it is said 'strong wind', but 'heavy rain'. It would be odd to say 'heavy wind' or 'strong rain'. Whilst all four of these words would be recognized by a learner at the pre-intermediate or even elementary level, it takes a greater degree of language competence to combine them correctly in productive use. To a native speaker, these combinations are highly predictable; to a learner, they are anything but. The full range of collocations is the following (Oxford Collocations Dictionary, 2002):

- **Adjective + noun:** bright/harsh/intense/strong light.
- **Quantifier + noun:** a beam/ray of light
- **Verb + noun:** cast/emit/give/provide/shed light.

- **Noun + verb:** light gleams/ glows/shines.
- **Noun + noun:** a light source.
- **Preposition + noun:** by the light of the moon.
- **Noun + preposition:** the light from the window.
- **Adverb + verb:** choose carefully.
- **Verb + verb:** be free to choose.
- **Verb + preposition:** choose between two things.
- **Verb + adjective:** make/ keep/ declare something safe.
- **Adverb + adjective:** perfectly/not entirely/environmentally safe.
- **Adjective + preposition:** safe from attack.

### ***Preposition***

A word that is used with nouns, pronouns, and gerunds to link them grammatically to other words. The phrase so formed, consisting of a preposition and its complement, is prepositional. In English, a prepositional phrase may be “discontinuous”, as in ‘Who(m) did you speak to?’ Prepositions may express such meanings as a possession (e.g., the leg of the table), direction (e.g., to the bank), place (e.g., at the corner), or time (e.g., before now) (Richards & Schmidt, 2002).

## **Literature Review**

### ***Studies in Corpus-based Strategies***

A corpus is a collection of texts in written or spoken forms stored on a computer and analyzed through computer software programs (O’Keefe et al., 2002). One of the main functions of a Corpus, as O’Keefe et al. (2007) pointed out is “to generate word frequency lists and concordance lines of words and clusters (or chunks)” (p. 1). This has led researchers to use corpora to deal with emergent phenomena regarding language analysis and teaching.

Johns (1991) was a pioneer for the use of corpora in language learning and highlighted their potential which has been widely acknowledged. With the emergence of ICTs, the use of corpora in language learning processes has become more common, giving students samples of language use in different contexts. Corpus researchers have revolutionized English language teaching as they have found that corpora usage fosters inductive learning (Kheirzadeh & Marandi, 2014) and improves learners’ language awareness (Biber et al., 1998; Granger, 1998; Johns, 1991; Kettemann, 1995).

Numerous studies have revealed that the use of corpora and Data-Driven Learning (DDL) assists students in solving writing problems and raising language awareness (Lee, 2011; Luo, 2015; Luo & Liao, 2015; Vyatkina, 2016; Yoon, 2008). Teachers have used corpora to provide students with opportunities to correct peers' writing and guide them to identify mistakes, give suggestions, and make accurate corrections. Students can benefit from the samples of language use found in a corpus to avoid lexicon-grammatical mistakes which can be translated into quality compositions. A corpus is likely to lead students to be more aware of their mistakes as they must analyze and compare language use samples before using them in free production practices.

Equally important, research studies have revealed how the implementation of corpora increased students' autonomy (Huang, 2011; Kheirzadeh & Marandi, 2014; Nesselhauf, 2004; Vyatkina, 2016; Yoon 2008; Zhang & Liu, 2014) and helped them build confidence in their L2 writing (Lai, 2015; Yoon & Hirvela, 2004), which are two aspects to highlight considering that a corpus is a database.

Hegelheimer's (2006) study focused on investigating the utility of an online resource to improve advanced-level ESL students' writing, finding a considerable error reduction in their writing while students exhibited a better understanding of language features when using a corpus-based resource called iWRITE. Other studies focused on using corpora as a tool for analyzing language patterns included Huang's (2008) which focused on analyzing the use of genitive (my) and accusative (me) subjects functioning as complements and examining meta-discourse. Reimerink (2006) analyzed the use of verbs in different sections of medical research articles, demonstrating, on one hand, that lexical domains were distributed differently in each part of an article, and on the other, that there was a complex relationship between a research article and its abstract.

Many studies have highlighted the difficulty of implementing a corpus in language teaching because of the training time, lack of analytical skills, and use of different commands to certain searches (Huang, 2011; Kheirzadeh & Marandi, 2014; Nesselhauf, 2004; Vyatkina, 2016; Yoon 2008; Yoon & Hirvela, 2004; Zhang & Liu, 2014). Other studies have shown benefits related to participants' positive attitudes toward corpus use in language teaching and learning processes (Braun, 2007; Gaskell & Cobb, 2004; Lee & Swales, 2006; Luo, 2016).

### ***Process Approach***

Writing develops higher-order skills which lead students to evaluate and analyze the information in-depth, develop unique points of view, and focus on using language conventions in contextually appropriate ways (Harmer, 2004). The process approach is recognized for its benefits and the teacher's role, whose main responsibility is to draw out students' potential, be a facilitator, and provide little or no input (Badger & White, 2000; Hyland, 2003). This

approach relates to linguistic skills (planning and drafting) rather than linguistic knowledge (grammar and text structures). Writing is mainly seen as a series of stages that, if undertaken during iterative writing, will develop language skills unconsciously.

Tribble (1996) identifies four stages to the process approach: prewriting, composing/drafting, revising, and editing (as cited in Badger & White, 2001). However, Hyland (2003) argues that this approach is recursive, interactive, and simultaneous as the writer can review, evaluate, revise, and rewrite parts of the work at any stage. Thus, Hyland (2003) urges teachers to avoid an emphasis on form since this approach aims at helping learners develop their strategies for generating, drafting, and refining ideas, and to mature their metacognitive writing.

Numerous studies have pointed out that the process approach is effective in different EFL contexts (Akinwamide, 2012; Alodwan & Ibnian, 2014; Ho, 2006; Laksmi, 2006). These studies demonstrated that students improved their writing production and became more autonomous. Similarly, other studies demonstrated that the process approach improved students' confidence and motivation as they were more secure about what and how they wanted to express themselves through their writing (Díaz Ramírez, 2014; Faraj, 2015; Melgarejo, 2010). Al-sawalha's (2014) study revealed the potential of process writing for English literature students. Before implementing the process approach, his participants did not feel at ease when writing; after implementing this approach, participants overcame writer's block, organized their ideas better, and gained the necessary cognitive skills to produce better writing.

### ***Studies in Writing Learning***

Numerous studies in Colombia have confirmed that advanced students at the university level, and even students of undergraduate foreign language teaching programs, have difficulties dealing with academic writing (Escobar Alméciga & Reid, 2014; Gómez, 2011; Robayo & Hernández, 2013; Zuñiga & Macias, 2006). Among the most common causes are poor writing instruction, difficulty with language structures, disorganization of ideas, and lack of discursive skills. Other studies have confirmed that the main difficulty of other EFL students (with an L1 other than Spanish) is related to the lack of previous experience and practice with the target language (Al-Badi, 2015; Sağlamel & Kayaoğlu, 2015).

Xudong et al.'s (2010) study focused on the benefits of explicit instruction and strategies to write in an academic context. This study demonstrated that the students who attended English for Specific Purposes courses improved their academic vocabulary and gained confidence in writing papers. Yamada (2000) found that summary writing activities used to teach textual patterns improved students' comprehension skills and writing production, and increased content knowledge in their academic fields. Castro Garcé's (2016) study

demonstrated that by demystifying academic writing and providing students with practice in doing different writing tasks, master's program students at a Colombian university could acknowledge the importance of producing academic texts in their professional lives and gain confidence to continue facing challenges regarding this skill.

## Methodology

### *Research Design*

This study had an action research design which involves identifying problematic situations in the classroom to find a solution by being reflective, discerning, and analyzing teaching situations in depth (Burns, 2010; Hopkins, 2008). Moreover, it was informed by a convergent mixed-method design (Creswell, 2014; Merriam & Tisdell, 2016) in which the quantitative and the qualitative data were collected concurrently, analyzed separately, and then compared to determine whether the findings of the data set converge with or diverge from each other (Creswell & Guetterman, 2021).

Both data sets were equally weighted by collecting information from the same number of individuals. This design was chosen “to increase the validity of construct and inquiry inferences” (Greene, 2007, p. 100). Using both data sets served to obtain more information rooted in the participants' transactions and interactions, offering a holistic understanding of the phenomenon under study. The qualitative strand of the study consisted of the responses to the open-ended questions in the pre- and post-implementation surveys and the teacher's journal notes analysis. The quantitative strand referred to analyzing the number of collocation and preposition mistakes in students' compositions.

### *Data Collection Methods*

Data were collected using three methods, surveys, a teacher's journal, and classroom artifacts. The participants completed three surveys (pre-, while-, and post-implementation). The surveys (Appendix A) had closed and open-ended questions to gather information about students' digital competence level, opinions regarding the use of ICT tools at school, previous experience with corpora for educational purposes, and perceptions regarding the whole implementation process.

The teacher's journal (Appendix B) was used *in situ* to record classroom situations and participants' reactions, attitudes, and behaviors toward corpus use. This instrument also aimed at gathering data about the lesson objectives, what went well, and what the teacher would have done differently in the lesson. Moreover, the journal included two spaces in which the researcher could write down overall descriptions of in-class situations and reflections on



the lesson. Entries to the journal were written immediately after each session to guarantee the accuracy of descriptions and thoughts.

Considering the nature of this study, the classroom artifacts (Appendix C) were essential because they allowed the researcher to identify students' difficulties and accomplishments, give them feedback, track their improvement process, and observe their outcomes after using the corpus for essay writing correction. As part of this instrument, the researcher used a correction code (Appendix D) to revise students' papers and treat their tasks with the same rigor.

### ***Participants***

14 tenth-grade students (N=14) from a bilingual private school in Bogota, Colombia, participated in this study. Their ages ranged from 15 to 17. They were selected using purposeful sampling which seeks "to discover, understand, and gain insight" (Merriam, 1998, p. 61) on a phenomenon. The participants were selected considering their English proficiency level. They had a B2 level according to standards established in the Common European Framework of Reference for Languages, CEFR (Council of Europe, 2014), which favored the purposes of the study as people who have this level can produce coherent and cohesive compositions.

### ***Ethical Considerations***

To ensure the ethical treatment of participants--all minor students-- consent forms were distributed to inform them and their parents about the study (Appendix E). The consent forms clarified that the study did not aim to evaluate their performance, collect personal information, or cause any harm. Additionally, participants were informed that their involvement was voluntary and that they could withdraw at any time without any consequences. These forms and all related documentation were reviewed and approved by the director of the Master's Program in English Language Teaching at a private university in Chia, Colombia.

All participants' information was anonymized by assigning each participant a unique code consisting of numbers and letters. The data sets were organized and stored in an encrypted folder on the researcher's laptop, ensuring data protection and preventing breaches.

### ***Implementation***

The pedagogical implementation was carried out in a nine-week period, which included a pre-implementation stage week, a four-week training stage, and a four-week production stage. In the pre-implementation week, the project was socialized with the school board, students, and students' parents. The researcher conducted a pre-implementation survey to gather information regarding students' previous knowledge about essay writing, process approach, and corpora.

Then, during the training stage, the students attended four input workshops related to the stages of the process approach (pre-writing, drafting, revising, and editing). Subsequently, the students attended workshops on essay writing structure, collocations and prepositions, correction code, and using the corpus (COCA). After this stage, a while implementation survey was conducted to assess the workshops, identify learners' difficulties during the training, and make the necessary modifications for the next stage. Finally, during the four-week production stage, learners focused on writing their essays about a given topic following the stages of the process approach, using the corpus in the editing stage of their writing process, and submitting a final version of their essays. As the researcher, I designed the action plan including the training stages, the instructional material, and data collection instruments.

## Findings

The qualitative data were analyzed using thematic analysis (Braun & Clarke, 2006). First, I conducted a preliminary data analysis (Grbich, 2013), which involved reviewing my data sets twice, taking notes of key points related to the topic of my research study, and memoing to keep track of emerging ideas during the preliminary analysis. This process favored the identification of concurrent patterns and the configuration of a set of codes. Then, I used an axial and selective coding strategy (Creswell, 2012), enabling me to classify and arrange several pieces of information, examine relationships in the data more efficiently, and generate the final themes. This process was guided by an inductive approach, avoiding the influence of a pre-existing coding framework and my analytical biases (Braun & Clarke, 2006).

A one-sample t-test, with an alpha level of 0.05 served to analyze the quantitative data, facilitating the identification of numerical trends and patterns of language use. The data were displayed through raw scores and percentages, graphs, and a descriptive analysis. A methodological triangulation was used to analyze data by comparing and merging qualitative and quantitative findings. To understand the findings completely, I selected the most relevant data to be detailed and organized into the identified themes.

## Enhancing Error Correction Autonomy and Language Awareness

### *Correcting Writing Errors Autonomously*

After analyzing the post-implementation surveys and the teacher's journal, the participants expressed they had benefitted from incorporating the corpus in various ways, more specifically gains associated with their autonomy when correcting their papers (See excerpt 1).

**Excerpt 1.** Responses to question five post-implementation survey, “Do you think that the use of a corpus has helped you become more autonomous when correcting writing mistakes? Yes - No Why?” (Unedited)

“Yes, because if I received a good feedback [sic], I could use it to correct my mistakes by myself” (Student 2)

“Yes, because I don’t need to take time from my break or a teacher who can tell me if I am writing or using collocations or prepositions well. On the contrary, I can correct my mistakes by myself using the tool” (Student 3)

“Yes, because I don’t have to ask for help to somebody else, I can correct my mistakes on my own. It helps to save time and effort” (Student 8)

Participants had various opinions regarding the tool. The most salient benefits were related to the fact that they could correct themselves without the teacher’s intervention and save time. Furthermore, they kept using the tool individually for other purposes, indicating gains in confidence when using the tool toward the end of the implementation.

Similarly, the participants showed evolution regarding their autonomy as they got more familiar with the tool. While in the first interactions with the tool they required help, later in the process they started making decisions and using it without the teacher’s intervention as depicted in excerpts 2 and 3.

**Excerpt 2.** Notes taken from teacher’s journal. Session 1 and 2 (edition and correction sessions) (Unedited)

The students received their feedback and seemed to understand the comments and suggestions provided by me as they started working immediately with the corpus. At the beginning of this editing and error correction sessions, some students (8 students) required my intervention to remind them the commands and help them make decisions regarding the results showed by the corpus. I helped them remember the commands, but I encouraged them to make a decision in relation with the search results. (Teacher’s journal comment)

**Excerpt 3.** Notes taken from teacher’s journal. Sessions 3 and 4 (edition and correction process) (Unedited)

At the beginning of the second editing and correcting sessions, the students were more confident regarding making decisions and how to interpret the results provided by the corpus. In fact, they did not require my intervention at all in comparison to the first session when they had to use the corpus. Just a few of them (2 students) asked me to remind them the commands. However, in terms of how to interpret results and decision-making, they managed to do it by themselves. (Teacher’s journal comment)

As seen in the excerpts, there is a great difference between the number of students who required help during the first two editing and correction sessions and the third and fourth sessions. A feasible explanation for this phenomenon is that the participants had never used a corpus under uncontrolled conditions before the first implementation. Instead, they used it under controlled conditions with constant teacher monitoring during the training stage. Hence, students gained more and more autonomy and confidence as they grew more familiar with the tool during the editing and writing correction sessions.

Likewise, the participants reported positive effects on their autonomy, documented in their opinions in excerpt 4 in which they had their points of view regarding the usefulness of the corpus. Many participants noticed that this tool helped them correct their mistakes and realize how to write their compositions better using examples.

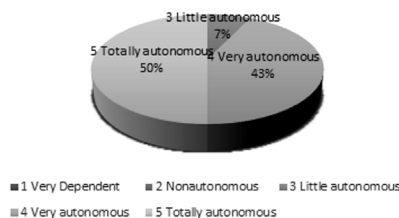
**Excerpt 4.** Post-implementation survey, question 2. “Do you think the corpus (COCA) is useful to correct writing mistakes? Yes – No, why?”

“Yes, because the collocation mistakes can be corrected using this tool. In fact, each time that I have a doubt about the use of a collocation or a verb in certain context, I use this tool” (Student 3)

“Yes, because it is sometimes difficult to realize that a word is used wrongly, and the corpus allows you to see how you can use them correctly” (Student 14)

Moreover, the participants reported becoming more autonomous toward the end of the implementation. Figure 1 depicts how they answered the following question “On a scale from 1 to 5, how autonomous have you become when correcting your compositions?” 50% of the participants felt “totally autonomous” and 43% “very autonomous,” depicting a positive trend regarding student autonomy, which grew more as they got familiar with the tool.

**How autonomous have you become when correcting your compositions?**



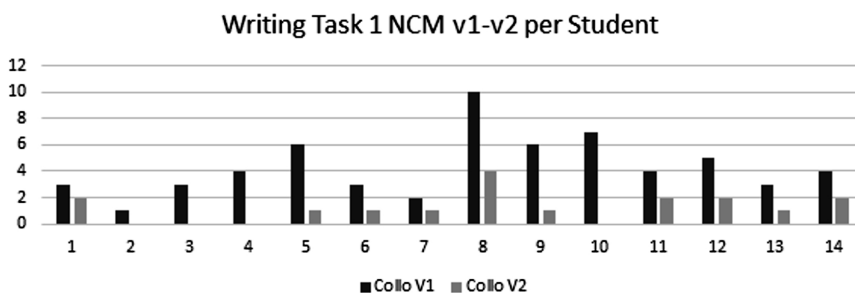
**Figure 1.** Percentages of Answers for the Question “How autonomous have you become when correcting your compositions?”

1. Very dependent, 2. Nonautonomous, 3. Little autonomous, 4. Very autonomous, 5. Totally autonomous.

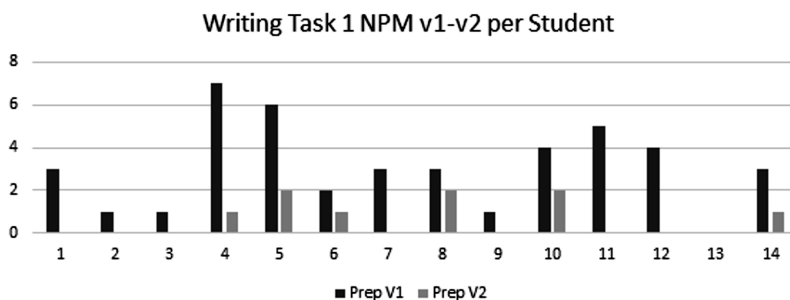
In brief, the participants reported gaining more autonomy when writing their compositions and correcting them using the corpus during and toward the end of the intervention. Although most of the data in this theme were withdrawn from the post-implementation survey, the analysis of the notes and comments on different phases of the intervention in the teacher’s journal, students’ compositions, and answers to open-ended questions confirmed this tool may be useful enhancing student autonomy in writing development processes. Moreover, the analysis also showed the participants’ perception shifts regarding using the corpus in different stages of the intervention, indicating an evolution process.

### *Improvements in Lexical, Grammatical, and Language Awareness Issues*

The bar charts (Figures 2 and 3) depict how the participants’ compositions evolved from one version to the other concerning the lexical items under study. These charts deal with the variation of error rates – collocations and prepositions - between the first version (dark grey) and the final version (light grey) of the writing task one.



**Figure 2.** This Figure Contrasts the Number of Collocation Mistakes (NCM) per Student in the First (dark grey) and Second Versions (light grey) of the Writing Task 1



**Figure 3.** This Figure Contrasts the Number of Preposition Mistakes (NPM) per Student in the First (dark grey) and Second Versions (light grey) of the Writing Task 1

Both graphs indicate that all the participants decreased their number of collocations and prepositions mistakes considerably from the first to the second version of writing task 1. In Figure 1, although there may not have been considerable improvement among all students regarding collocation mistakes, every student showed a moderate amelioration. Indeed, only 28% of them presented an improved final version of the essay. As for preposition mistakes, there was a similar trend, but in this specific case, the improvement was more significant as 43% of the students submitted the final version of their essay with no errors of this nature.

Apart from the preliminary analysis, findings show the results of a paired t-test, which provides more detailed information regarding a statistically significant difference. As can be gleaned from Table 1, the data needed for this test were the number of collocation mistakes (NCM) and number of preposition mistakes (NPM) in writing task 1, during Pre-test (before using a corpus, V1) and Post-test (after integrating a corpus, V2). For the significance level,  $\alpha = 0.05$  is used. The Null Hypothesis will be rejected if the p-value is less than ( $<$ ) 0.05. On the other hand, if it is more than ( $>$ ) 0.05, the Null Hypothesis will be accepted.

### Hypothesis one:

- $H_0$ : Implementing a corpus-based strategy will increase the mistakes between the first and second versions of writing task 1.
- $H_1$ : Implementing a corpus-based strategy will decrease the mistakes between the first and second versions of writing task 1.

**Table 1.** *Pair Sample Test of Number of Collocations and Prepositions Mistakes in Writing Task One*

Paired Samples Test				
		t	df	Significance
		Two-Sided p		
Pair 1	NCM V1 NCM V2	6,013	13	<,001
Pair 2	NPM V1 NPM V2	5,090	13	<,001

*Note.* NCM = Number of Collocations Mistakes; NPM = Number of Prepositions Mistakes; V= version.

From the statistical analysis using a paired sample t-test, it was apparent that the p-value for the Pre-test and Post-test from both items NCM and NPM respectively were 0,001 and

0,001. This result means that all the p-values are less than 0.05. Thus, the difference in the results is statistically significant.

These findings were confirmed by comparing the *t*-value and the Critical Value (CV). The *t* value has to be larger than the CV to reject the Null Hypothesis. The Degrees of Freedom (*df*) and the T-distribution table were used to determine the CV. The reference in the table was the cumulative probability of 90,  $\text{com.prob} = t_{.90}$ , and two-tails. The CV value was equal to 1.350 and the *t* value for the Pre-test and Post-test from both items NCM and NPM respectively were 6,013 and 5,090, indicating that the *t* values are much larger than the CV. In addition, the T-distribution table indicates that the results had an 80% of confidence level, confirming that the Null Hypothesis can be rejected and the results are statistically significant.

The same paired-sample test was used to determine whether there was a difference in the average number of collocation and preposition mistakes between the Pre-test and the Post-test in terms of NCM and NPM.

### Hypothesis two:

- $H_0$ : The average of collocation and preposition mistakes of the Pre-test is equal to the Post-test ( $\mu_{\text{Pre}} = \mu_{\text{Post}}$ ).
- $H_1$ : The average of collocation and preposition mistakes of the Pre-test is not equal to the Post-test ( $\mu_{\text{Pre}} \neq \mu_{\text{Post}}$ ).

Table 2 shows that the average number of collocation and preposition mistakes decreased considerably between the first and second versions of writing task 1; this went from 4.3 mistakes on average to 1.2 in the whole sample. Similarly, the average number of preposition mistakes decreased substantially, going from 3 to 0.64.

**Table 2.** *Mean of Collocations and Prepositions Mistakes*

Paired Samples Statistics				
		Mean	N	Std. Deviation
Pair 1	Number of Collocations Mistakes Version 1	4,3571	14	2,30742
	Number of Collocations Mistakes Version 2	1,2143	14	1,12171
Pair 2	Number of Prepositions Mistakes Version 1	3,0714	14	2,01778
	Number of Prepositions Mistakes Version 2	,6429	14	,84190

Considering the abovementioned, implementing a corpus for writing error correction shows 100% effectiveness as all the participants decreased the number of mistakes in different proportions. More importantly, four participants of the total sample submitted the final version without collocation mistakes. As for preposition mistakes, the relation was eight out of fourteen who submitted the final version without them.

As for improvements in grammatical, language awareness, and lexical issues, many students maintained that the corpus had made them more aware of the importance of lexical and semantical issues (See excerpt 5).

**Excerpt 5.** Post-implementation survey (unedited).

*“Yes, because it is a tool that helps you to comprehend more easily phrases or words by using examples” (Student 4)*

*“Yes, it helps you to find new words and do not repeat the ones that you already use. Besides, it gives you ideas on how to use a word in different contexts” (Student 8)*

*“The corpus helped me with word order, and understand the use of words in different contexts” (Student 11)*

*“it makes more aware of your mistakes” (Student 9)*

*“I think more what would be the best way to express my ideas” (Student 12)*

Reflecting on the above, although the students had difficulty using the corpus during the first editing and error correction session, they overcame the difficulties and benefited from the corpus used for writing correction processes. Students’ opinion also served to corroborate the gains of using a corpus; most of them pointed out that the tool had assisted them on understanding how to use words in specific contexts and improving the quality of their ideas.

## Shortcomings of Using a Corpus in Writing Error Correction

104

### *Recalling Commands from the Interface and Associating Them with Linguistic Categories*

The findings revealed that the participants had difficulties using the corpus. Certainly, when the participants were asked about the main difficulties of using a corpus, they mentioned that the commands (combinations of words to search) were one of them, which is evident in excerpt 6.



**Excerpt 6.** Post-implementation survey (unedited).

*"I had some difficulty when I used the corpus because I could not remember the commands easily" (Student 3)*

*"You need the commands that are difficult to remember, and sometimes it is difficult to identify what command is the most appropriate" (Student 14)*

*"I was confused with the use of some commands" (Student 8)*

As seen, the students complained about difficulty remembering the commands. Five out of fourteen students claimed to have difficulties doing so. This was corroborated by the notes in the teacher's journal (Excerpt 7).

**Excerpt 7.** Teacher's journal comments. Production stage.

*"At the beginning of this editing and correcting session, some students required my intervention to remind them the commands" (teacher's journal)*

*"Just a few of them asked me to remind them the commands" (teacher's journal)*

However, this difficulty was not due to the participants' bad memory as the teacher taught just four commands and put them into practice during the training stage. Their difficulty was more attributed to their lack of language awareness about linguistic categories. See excerpt 8.

**Excerpt 8.** Teacher's journal comments. Training stage.

*"Most of the students had difficulty to use the commands because they could not differentiate some lexical categories (for example: adverbs and adjectives) and each of those categories have [sic] an assigned command in the interface" (teacher's journal)*

Another reason to explain why the participants had difficulty using the commands properly is related to the fact that they could not easily associate the lexical categories with their corresponding representations in the corpus. This situation is evident in the excerpts 9 and 10.

**Excerpt 9.** Teacher's journal comments. Training stage.

*"Most of the students had difficulty to use [sic] the commands because they could not differentiate some lexical categories, nor to associate them easily to the representations that the linguistic category had in the corpus. For instance, in the corpus an adverb is represented by a "r" and adjectives by a "j". Thus, they did not understand and questioned why those commands were represented by those letters." (teacher's journal)*

**Excerpt 10.** Post-implementation survey, question 4, On a scale from 1 to 5, how difficult was it to use a corpus (COCA) for correcting writing mistakes? Being 1 Not difficult and 5 Too difficult. Why?

*“It still creates some difficulty. I had some difficulty when I used the corpus because I could not remember the commands easily. However, if you have them on hand, this tool makes your life easier” (Student 3)*

*“It is not difficult, because the use of this tool is very simple, it is necessary to learn the commands by memory or have them noted to find the options” (Student 9)*

Toward the end of the implementation, the students could use the commands well by making lists of them and the linguistic categories. Briefly, although a corpus for writing error correction could benefit students positively in different aspects, including language awareness and autonomy, this also might represent difficulties for students with poor language awareness. Thus, it is critical to consider what difficulties emerge using a corpus in EFL scenarios so that the training stage design goes properly based on participants’ needs and characteristics.

### *Lack of Self-efficacy in Making Decisions*

One of the problematic situations when using a corpus for error correction is related to the interpretation of the high number of results. When a person searches for a possible combination, the corpus normally displays numerous combinations and samples for each combination, as seen in Figures 4 and 5.

106

SEARCH	FREQUENCY	CONTEXT	ACCOUNT
81			
82			
83			
84			
85			
86			
87			
88			
89			
90			
91			
92			
93			
94			
95			
96			
97			
98			
99			
100			
TOTAL			17794

Figure 4. Example Output of a Search (Do\_n) in the Interface

In Figure 4, it is depicted a search of the command “Do\_n” which requests the corpus of finding collocations of the verb “do” with potential nouns, resulting in 100 possible combinations of “do” with nouns and 17,394 examples in total. These results hindered students’ decision-making and made them feel overwhelmed. Similarly, Figure 5 displays a search of the command “think\_i” which requests the corpus of finding collocations of the verb “think” with potential prepositions, resulting in 85 possible combinations and 85,448 examples.

Corpus of Contemporary American English			
SEARCH	FREQUENCY	CONTEXT	ACCOUNT
67	<input type="checkbox"/>	THINK WORTH	2
68	<input type="checkbox"/>	THINK VICE	1
69	<input type="checkbox"/>	THINK VERSUS	1
70	<input type="checkbox"/>	THINK POST	1
71	<input type="checkbox"/>	THINK SAVE	1
72	<input type="checkbox"/>	THINK RELATIVE	1
73	<input type="checkbox"/>	THINK OUT-SIDE	1
74	<input type="checkbox"/>	THINK ONTO	1
75	<input type="checkbox"/>	THINK O'ER	1
76	<input type="checkbox"/>	THINK NOTWITHSTANDING	1
77	<input type="checkbox"/>	THINK LET	1
78	<input type="checkbox"/>	THINK EXCLUDING	1
79	<input type="checkbox"/>	THINK CONTRARY	1
80	<input type="checkbox"/>	THINK CONSIDERING	1
81	<input type="checkbox"/>	THINK BUT	1
82	<input type="checkbox"/>	THINK BENEATH	1
83	<input type="checkbox"/>	THINK BELOW	1
84	<input type="checkbox"/>	THINK BARRING	1
85	<input type="checkbox"/>	THINK ALIKE	1
TOTAL			85,448

Figure 5. Example Output of a Search (think\_i) in the Interface

This issue is another difficulty the students faced, as they had to decide on the most suitable option for their writing task by analyzing samples. This process was time-consuming and quite difficult for students with poor analytical skills. Furthermore, I confirmed students’ difficulty with the number of results per search after analyzing post-implementation survey answers and notes in the teacher’s journal, as evident in excerpts 11 and 12.

**Excerpt 11.** Post-implementation survey, question 4, On a scale from 1 to 5, how difficult was it to use a corpus (COCA) for correcting writing mistakes? Being 1 Not difficult and 5 Too difficult. Why?

*“It still creates some difficulty. Sometimes it was difficult to take a decision about what option given by the corpus to use” (Student 8)*

**Excerpt 12.** Teacher’s journal comments.

*“I think they needed more practice on how to use the corpus and results interpretation” (teacher’s journal)*

*“I will make a tutorial video to reinforce their (participants) knowledge about how to use the corpus and how to interpret results so that they can feel more confident when using the tool. (teacher’s journal)*

Although the participants had difficulty making decisions and interpreting the results, additional training and support materials were provided to students for using the corpus effectively through the training and production stages. Moreover, using a corpus for error correction in writing may create difficulties of a different nature. Nevertheless, the findings revealed that by implementing a corpus-based strategy appropriately, the participants may revise their essays autonomously before producing a final version and enhancing their language awareness. More importantly, the participants understood what was involved in producing high-quality and error-free written compositions.

## Discussion

The present study showed the importance of incorporating error correction processes in EFL scenarios to further student writing production improvements. First, the students could self-correct their collocation mistakes using a corpus as a language use reference, relevant as teaching about collocation use has been neglected in ELT classrooms (Hashemi et al., 2012). Similarly, the corpus contributed to deal with the use of prepositions, which is a difficult aspect to master due to their polysemous nature (Lorincz & Gordon, 2012). This is aligned with previous studies that found positive outcomes related to error correction using a corpus (Gaskell & Cobb, 2004; Lai, 2015; Smirnova, 2017; Tono et al., 2014; Yoon & Jo, 2014). Most of the students of this study showed improvements over the production and edition stages and found the tool beneficial and functional. Although a few of them had difficulty using the tool at the beginning of the process, this was addressed by revisiting some concepts to understand the tool commands and providing them with more guided practice.

In addition, the students demonstrably improved their language awareness, confirming the outcomes of previous studies (Hegelheimer, 2006; Phoocharoensil, 2012; Vieira, 2013; Vyatkina, 2016). The students could differentiate parts of the speech (verbs, nouns, adjectives, adverbs, among others) and learn to use them properly in written tasks by the end of the implementation. Although this was out of the scope of the study, an emergent theme in the data analysis showed that the participants indirectly improved their language awareness, particularly with the correct use of the corpus.

Implementing corpora in error correction processes fosters students' autonomy considerably echoing the conclusions of other researchers (Chambers, 2005; Huang, 2011; Kheirzadeh & Marandi, 2014; Luo & Liao, 2015; Yoon, 2008; Zhang & Liu, 2014). Over 80% of the participants expressed that, toward the end of the intervention, they needed less intervention from the teacher; they kept using the tool for corroborating the correct use of the lexicon. This was aligned with the notes of the teacher's journal which indicated that, as weeks went by, the students requested less help.

Finally, students' successful use of corpora for error correction requires well-planned training and practice. Although this study did not aim to find issues related to the constraints, training is key to achieving the expected outcomes. This validates the findings of a previous study that concluded that training is an important factor when using corpora in the classroom (Vannestål & Lindquist, 2007). Additionally, the findings of this research pitted this researcher against Boulton's (2010) conclusion, which asserted that implementing corpora in teaching might be too technically challenging.

Regardless of the difficulties that using a corpus may pose to educators in language teaching, technology-based tools are unstoppable, and students are using them more. Cellphones, tablets, social networks, and Artificial Intelligence draw students' attention and keep them on track through exposure to fast content and solutions. This situation demands teachers and school administrators adaptative dynamics to provide students with more meaningful content based on their needs and what new technologies impose. This must serve to advocate for the implementation of more robust training programs for educators and policy on how to integrate technology purposefully in educational settings.

Additionally, the role of teachers and students demands a shift in the dynamics in the classroom. Ultimately, students have already realized that they can learn at their own pace and with the material of their preference. This requires teachers to redesign their lessons carefully using instructional design models and technologies that favor their learning environment. This, in turn, will allow teachers to focus more on integrating strategies that empower and encourage learners to have more agency in their learning process.

## Limitations of the Study

One of the major limitations of this research was the participants' lack of experience in writing correction procedures. They did not initially know how to organize their ideas or draw an outline. Moreover, they were not familiar with correction codes or annotated writing tasks. This limitation posed difficulty as I had to invest more time than expected in introducing the participants to various procedures and concepts so that they navigated smoothly throughout the implementation phases.

Another limitation was related to the time available for the intervention. It was impossible to initiate the intervention as planned because the school board allocated just a few hours of the total English lessons per week. Thus, the researcher modified and planned new activities and materials to provide students with optional input. For this purpose, the researcher made video tutorials, worksheets, and PowerPoint presentations to cover all the contents necessary for a successful intervention and favoring the student learning curve.

## References

- Akinwamide, T. K. (2012). The influence of process approach on English as second language students' performances in essay writing. *English Language Teaching*, 5(3), 16–29. <https://doi.org/10.5539/elt.v5n3p16>
- Al-Badi, I. A. H. (2015). Academic writing difficulties of ESL learners. *The 2015 WEI International Academic Conference Proceedings* (pp. 65–78) <http://www.westeastinstitute.com/wp-content/uploads/2015/02/Ibtisam-Ali-Hassan-Al-Badi-full-Paper.pdf>
- Al-sawalha, A. M. (2014). Potential advantages of process writing for students English language and literature at Jerash University in Jordan DR. *International Journal on Studies in English Language and Literature (IJSELL)*, 2(6), 41–46. <https://www.arcjournals.org/pdfs/ijSELL/v2-i6/6.pdf>
- Alodwan, T. A. A., & Ibnian, S. S. K. (2014). The effect of using the process approach to writing on developing university students' essay writing skills in EFL. *International Journal of Linguistics and Communication*, 2(2), 147–163. [https://ijlc.thebrpi.org/journals/ijlc/Vol\\_2\\_No\\_2\\_June\\_2014/9.pdf](https://ijlc.thebrpi.org/journals/ijlc/Vol_2_No_2_June_2014/9.pdf)
- Badger, R., & White, G. (2000). A process genre approach to teaching writing. *ELT Journal*, 54(2), 153–160. <https://doi.org/10.1093/elt/54.2.153>
- Benko, S. L. (2012). *Teaching to the task: Preservice teachers' instruction for cognitively demanding writing tasks* (Publication No. 3532772) [Doctoral Dissertation, University of Pittsburgh]. Available from ProQuest Dissertations & Theses Global. <https://proxyiub.uits.iu.edu/login?url=https%3A%2F%2Fwww.proquest.com%2Fdissertations-theses%2Fteaching-task-preservice-teachers-instruction%2Fdocview%2F1222084061%2Fse-2%3Faccountid%3D11620>
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge University Press.
- Boulton, A. (2010). Data-driven learning: Taking the computer out of the equation. *Language Learning*, 60(3), 534–572. <https://doi.org/10.1111/j.1467-9922.2010.00566.x>
- Braun, S. (2007). Integrating corpus work into secondary education: From data-driven learning to needs-driven corpora. *ReCALL*, 19(3), 307–328. <https://doi.org/10.1017/S0958344007000535>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/proxyiub.uits.iu.edu/10.1191/1478088706qp063oa>
- Brown, H. D. (2007). *Principles of language learning and teaching* (5th ed.). Pearson Longman
- Burns, A. (2010). *Doing action research in English language teaching. A guide for practitioners*. Taylor & Francis.
- Castro Garcé, A. Y. (2016). Academic writing at graduate school: An attempt at its demystification. *World Journal of English Language*, 6(1), 25–34. <https://doi.org/10.5430/wjel.v6n1p25>
- Chambers, A. (2005). Integrating corpus consultation in language studies. *Language Learning & Technology*, 9(2), 111–125. <http://dx.doi.org/10125/44022>

- Council for Cultural Co-operation Modern Languages Division. (2011). *Common European framework of reference for languages: Learning, teaching, assessment* (12. print). Cambridge University Press.
- Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Pearson Education, Inc.
- Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approach*. Sage Publications.
- Creswell, J.W., & Guetterman, T.C. (2021). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. Pearson Education Limited.
- Díaz Ramírez, M. (2014). The impact of project work and the writing process method on writing production. *HOW*, 21(2), 31–53. <https://doi.org/10.19183/how.21.2.3>
- Dudeny, G., & Hockly, N. (2007). *How to teach English with technology*. Pearson Longman.
- Escobar Alméciga, W. Y., & Reid, E. (2014). Mentor texts and the coding of academic writing structures: A functional approach. *HOW*, 21(2), 94–111. <https://doi.org/10.19183/how.21.2.6>
- Faraj, A. K. A. (2015). Scaffolding EFL students' writing through the writing process approach. *Journal of Education and Practice*, 6(13), 131–141. <https://iiste.org/Journals/index.php/JEP/article/view/22371/23017>
- Gaskell, D., & Cobb, T. (2004). Can learners use concordance feedback for writing errors? *System*, 32(3), 301–319. <https://doi.org/10.1016/j.system.2004.04.001>
- Gómez, J. D. (2011). Teaching EFL academic writing in Colombia: Reflections in contrastive rhetoric. *Profile: Issues in Teachers' Professional Development*, 13(1), 205–213. <http://www.revistas.unal.edu.co/index.php/profile/article/view/20577>
- Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology*, 99(3), 445–476. <https://doi.org/10.1037/0022-0663.99.3.445>
- Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A. P. Conie (Ed.), *Phraseology: Theory, analysis and applications*. (pp. 145–160). Oxford Academic. <https://doi.org/10.1093/oso/9780198294252.003.007>
- Grbich, C. (2013). *Qualitative data analysis: An introduction*. SAGE Publications Ltd.
- Greene, J. C. (2007). *Mixed methods in social inquiry*. Jossey-Bass.
- Harmer, J. (2004). *How to teach writing*. Longman.
- Hashemi, M., Azizinezhad, M., & Dravishi, S. (2012). Collocation a neglected aspect in teaching and learning EFL. *Procedia - Social and Behavioral Sciences*, 31, 522–525. <https://doi.org/10.1016/j.sbspro.2011.12.097>
- Hegelheimer, V. (2006). Helping ESL writers through a multimodal, corpus-based, online grammar resource. *CALICO Journal*, 24(1), 5–32. <https://www.jstor.org/stable/24156292>

- Ho, B. (2006). Using the process approach to teach writing in six Hong Kong primary classrooms. *New Horizons in Education*, 53, 22-41. <http://www.hkta1934.org.hk/NewHorizon/abstract/2006m/22-41.pdf>
- Hoff, E., & Core, C. (2013). Input and language development in bilingually developing children. *Seminars in Speech and Language*, 34(4), 215–226. <https://doi.org/10.1055/s-0033-1353448>
- Hopkins, D. (2008). *A teacher's guide to classroom research*. Open University Press.
- Huang, L. S. (2008). Using guided, corpus-aided discovery to generate active learning. *English Teaching Forum*, 46(4) 20–27. [https://americanenglish.state.gov/files/ae/resource\\_files/08-46-4-d.pdf](https://americanenglish.state.gov/files/ae/resource_files/08-46-4-d.pdf)
- Huang, L. S. (2011). Corpus-aided language learning. *ELT Journal*, 65(4), 481–484. <https://doi.org/10.1093/elt/ccr031>
- Hyland, K. (2002). *Teaching and researching writing*. Longman.
- Hyland, K. (2003). *Second language writing*. Cambridge University Press.
- Johns, T. (1991). Should you be persuaded: Two examples of data driven learning. *ELR Journal*, (4), 1–16. [http://www.lexically.net/wordsmith/corpus\\_linguistics\\_links/Tim Johns and DDL.pdf](http://www.lexically.net/wordsmith/corpus_linguistics_links/Tim%20Johns%20and%20DDL.pdf)
- Kennison, S. M. (2014). *Introduction to language development*. SAGE.
- Kettemann, B. (1995). On the use of concordancing in ELT. *AAA: Arbeiten Aus Anglistik Und Amerikanistik*, 20(1), 29–41. <http://www.jstor.org/stable/43023697>
- Kheirzadeh, S., & Marandi, S. S. (2014). Concordancing as a tool in learning collocations: The case of Iranian EFL learners. *Procedia - Social and Behavioral Sciences*, 98, 940–949. <https://doi.org/10.1016/j.sbspro.2014.03.503>
- Lai, S. L. (2015). EFL students' perceptions of corpus-tools as writing references. In Helm, Francesca; Bradley, Linda; Guarda, Marta; Thouësny, Sylvie (Eds), *Critical CALL – Proceedings of the 2015 EUROCALL Conference*, Padova, Italy (pp. 336-341). [Research-publishing.net. https://doi.org/10.14705/rpnet.2015.000355](https://doi.org/10.14705/rpnet.2015.000355)
- Laksmi, E. D. (2006). Scaffolding students writing in EFL class: Implementing process approach. *TEFLIN Journal*, 17(2), 144–156. <https://doi.org/10.15639/teflinjournal.v17i2/152-165>
- Lee, H. C. (2011). In defense of concordancing: An application of data-driven learning in Taiwan. *Procedia - Social and Behavioral Sciences*, 12, 399–408. <https://doi.org/10.1016/j.sbspro.2011.02.049>
- Lee, D., & Swales, J. (2006). A corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora. *English for Specific Purposes*, 25(1), 56–75. <https://doi.org/10.1016/j.esp.2005.02.010>
- Listyani, L. (2018). Enhancing academic writing skills through “Reading to Learn” strategy. *Arab World English Journal*, 9(1), 268–280. <https://doi.org/10.24093/awej/vol9no1.19>
- Lorincz, K., & Gordon, R. (2012). Difficulties in learning prepositions and possible solutions. *Linguistic Portfolios*, 1(1), 1–5. [https://repository.stcloudstate.edu/stcloud\\_ling/vol1/iss1/14](https://repository.stcloudstate.edu/stcloud_ling/vol1/iss1/14)



- Luo, Q. (2015). Application of data-driven learning in EFL writing instructions for engineering students. *World Transactions on Engineering and Technology Education*, 13(3), 238–244. [http://www.wiete.com.au/journals/WTE&TE/Pages/Vol.13,%20No.3%20\(2015\)/04-Luo-Q.pdf](http://www.wiete.com.au/journals/WTE&TE/Pages/Vol.13,%20No.3%20(2015)/04-Luo-Q.pdf)
- Luo, Q. (2016). The effects of data-driven learning activities on EFL learners' writing development. *Springer Plus*, 5(1255), 1–13. <https://doi.org/10.1186/s40064-016-2935-5>
- Luo, Q., & Liao, Y. (2015). Using corpora for error correction in EFL learners' writing. *Journal of Language Teaching and Research*, 6(6), 1333–1342. <http://dx.doi.org/10.17507/jltr.0606.22>
- Melgarejo, D. A. (2010). Assessing children's perceptions of writing in EFL based on the process approach. *Colombian Applied Linguistics Journal*, 12(1), 70–84. <https://doi.org/10.14483/22487085.92>
- Merriam, S. B. (1998). *Qualitative research and case study applications in education*. Jossey-Bass Publishers.
- Merriam, S. B., & Tisdell, E.J. (2016). *Qualitative Research. A Guide to Design and Implementation*. Jossey-Bass.
- Myles, J. (2002). Second language writing and research: The writing process and error analysis in student texts. *The Electronic Journal for English as a Second Language TESL-EJ*, 6(2), 1–19. <https://www.tesl-ej.org/cj22/a1.html>
- O'Keefe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge University Press.
- Oshima, A., & Hogue, A. (2007). *Introduction to academic writing. Level 3* (3<sup>rd</sup> ed.). Pearson Longman.
- Oxford collocations dictionary: for students of English. (2002). Oxford University Press.
- Phoocharoensil, S. (2012). Language corpora for EFL teachers: An exploration of English grammar through concordance lines. *Procedia - Social and Behavioral Sciences*, 64, 507–514. <https://doi.org/10.1016/j.sbspro.2012.11.060>
- Reimerink, A. (2006). The use of verbs in research articles: Corpus analysis for scientific writing and translation. *New Voices in Translation Studies*, 2, 9–27. <https://www.iatis.org/index.php/new-voices-in-translation-studies/item/325-issue-2-2006>
- Richards, J. C., & Schmidt, R. (2002). *Dictionary of language teaching and applied linguistics* (3<sup>rd</sup> ed.). Longman.
- Robayo, A., & Hernandez, L. (2013). Collaborative writing to enhance academic writing development through project work. *HOW*, 20(1), 130–148. <https://howjournalcolombia.org/index.php/how/article/view/27>
- Sağlamel, H., & Kayaoğlu, M. N. (2015). English major students' perceptions of academic writing: A struggle between writing to learn and learning to write. *Journal of History Culture and Art Research*, 4(3), 37. <https://doi.org/10.7596/taksad.v4i3.477>
- Smirnova, E. A. (2017). Using corpora in EFL classroom: The case study of IELTS preparation. *RELC Journal*, 48(3), 302–310. <https://doi.org/10.1177/0033688216684280>
- Stern, H. H. (1970). *Perspectives on second language teaching*. Ontario Institute for Studies in Education.

- Swales, J. M., & Feak, C. B. (2004). *Academic writing for graduate students: Essential tasks and skills*. University of Michigan Press.
- Tono, Y., Satake, Y., & Miura, A. (2014). The effects of using corpora on revision tasks in L2 writing with coded error feedback. *ReCALL*, 26(2), 147–162. <https://doi.org/10.1017/S095834401400007X>
- Tribble, C. (1996). *Writing*. Oxford University Press.
- Vannestål, M., & Lindquist, H. (2007). Learning English grammar with a corpus: Experimenting with concordancing in a university grammar course. *ReCALL*, 19(3), 329–350. <https://doi.org/10.1017/S0958344007000638>
- Vieira, N. G. S. (2013). Not to teach but give insights: Corpus-based approach in Portuguese-English and Portuguese-Russian cross-linguistic error correction. *Procedia - Social and Behavioral Sciences*, 95, 522–527. <http://linkinghub.elsevier.com/retrieve/pii/S1877042813041979>
- Vyatkina, N. (2016). Data-driven learning of collocations: Learner performance, proficiency, and perceptions. *Language Learning & Technology*, 20(3), 159–179. <http://dx.doi.org/10.125/44487>
- Xudong, D., Cheng, L., Varaprasad, C., & Meng, L. L. (2010). Academic writing development of ESL/EFL graduate students in NUS. *Reflections on English Language Teaching*, 9(2), 119–138. [https://www.nus.edu.sg/celc/wp-content/uploads/2022/11/119to138\\_deng.pdf](https://www.nus.edu.sg/celc/wp-content/uploads/2022/11/119to138_deng.pdf)
- Yamada, K. (2000). Perspectives: Helping novice EFL/ESL academic writers appreciate English textual patterns through summary writing. *JALT Journal*, 22(1), 196–208. <https://doi.org/10.37546/JALTJ22.1-10>
- Yoon, H. (2008). More than a linguistic reference: The influence of corpus technology on L2 academic writing. *Language Learning & Technology*, 12(2), 31–48. <http://dx.doi.org/10.125/44142>
- Yoon, H., & Hirvela, A. (2004). ESL student attitudes toward corpus use in L2 writing. *Journal of Second Language Writing*, 13(4), 257–283. <https://doi.org/10.1016/j.jslw.2004.06.002>
- Yoon, H., & Jo, J. W. (2014). Direct and indirect access to corpora: An exploratory case study comparing students' error correction and learning strategy use in L2 writing. *Language Learning & Technology*, 18(1). <http://dx.doi.org/10.125/44356>
- Zhang, Y., & Liu, L. (2014). A corpus-aided approach in EFL instruction: A case study of Chinese EFL learners' use of the infinitive. *English Language Teaching*, 7(7), 152–158. <https://doi.org/10.5539/elt.v7n7p152>
- Zúñiga, G., & Macias, D. F. (2006). Refining students' academic writing skills in an undergraduate foreign language teaching program. *Íkala, Revista de Lenguaje y Cultura* 11(1), 311–336. <https://doi.org/10.17533/udea.ikala.2792>

## Appendix A. Surveys

### Name of the Project: Configuring Collocations and Prepositions in Essay Writing through a Corpus-based Strategy

#### Pre-Implementation Survey

First, I would like to thank you for being part of this research study and inform you that this survey will be used for research purposes. Please, answer the questions herein marking with an (x) and justifying when necessary. Second, your answers will not be subject of any evaluation, therefore it is not mandatory to provide your full name.

Objective: This survey is aimed to identify your knowledge related to writing methodologies and computer-based tools.

**Name:** \_\_\_\_\_ **Date:** \_\_\_\_\_

1. Do you know any writing methodology (Specific steps to do a writing task)? If your answer is Yes, write which one.

Yes \_\_\_\_\_

No \_\_\_\_\_

**Which one?** \_\_\_\_\_

2. How useful may be using a writing methodology when doing your writing assignments? Explain

\_\_\_\_\_  
\_\_\_\_\_

3. Do you think that computer-based tools (computerized instruments or software) could contribute with your writing process?

Yes \_\_\_\_\_

No \_\_\_\_\_

Why?

\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

4. Have you ever used computer-based tools such as online dictionaries, translators, tutorial videos, etc.; to improve the quality of your writing tasks?  
Yes \_\_\_\_\_ No \_\_\_\_\_

**If YES, Which ones? (You can select more than 1 option)**

- a. Online Dictionaries
  - b. Forums
  - c. Translators
  - d. Tutorial videos
  - e. Blogs
  - f. Corpora (Data bases that contain examples of the use of a language)
  - g. Wikis
  - h. Social Network
5. How long do you use the abovementioned tools to do your writing tasks?
- a. Less than 1 hour per week
  - b. 1-2 hours per week
  - c. 3-6 hours per week
  - d. More than 6 hours per week

## Name of the Project: Configuring Collocations and Prepositions in Essay Writing through a Corpus-based Strategy

### While-Implementation Survey

First, I would like to thank you for being part of this research study and inform you that this survey will be used for research purposes. Please, answer the questions herein marking with an (x) and justifying when necessary. Second, your answers will not be subject of any evaluation, therefore it is not mandatory to provide your full name.

Name: \_\_\_\_\_ Date: \_\_\_\_\_

**1. How useful have you found the training workshops? Mark only one number for each item**

Workshops	1 Very useless	2 Useless	3 Fairly useful	4 Useful	5 Very useful
Process Approach					
Essay Writing Structure and Correction Code					
Collocations and Prepositions					
How corpora works					

**2. Which workshop did you find the most difficult?**

- Collocations and prepositions
- Essay writing structure and correction code
- Process approach
- How corpora works

Why? \_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

**3. What recommendations would you like to make about the given workshops?**

\_\_\_\_\_  
\_\_\_\_\_

## Name of the Project: Configuring Collocations and Prepositions in Essay Writing through a Corpus-based Strategy

### Post-Implementation Survey

First, I would like to thank you for being part of this research study and inform you that this survey will be used for research purposes. Please, answer the questions herein marking with an (x) and justifying when necessary. Second, your answers will not be subject of any evaluation, therefore it is not mandatory to provide your full name.

Name: \_\_\_\_\_ Date: \_\_\_\_\_

1. After using the process approach (pre-writing, drafting, editing and publishing), do you think you have improved your writing skills?

Yes \_\_\_\_\_

No \_\_\_\_\_

Why? \_\_\_\_\_

\_\_\_\_\_

How? \_\_\_\_\_

Why not? \_\_\_\_\_

2. Do you think the corpus (COCA) is useful to correct writing mistakes?

Yes \_\_\_\_\_

No \_\_\_\_\_

Why? \_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

118

3. In which aspects do you consider the corpus contributed the most during the editing stage? (You can choose more than one option)

a. Use of prepositions

b. Collocations

c. Spelling

d. Word order

e. None of them

4. On a scale from 1 to 5, how difficult was it to use a corpus (COCA) for correcting writing mistakes?

Not difficult (1) Little difficult (2) Fairly difficult (3) Difficult (4) Too difficult (5)

Why? \_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

5. Do you think that the use of a corpus has helped you to become more autonomous when correcting writing mistakes?

Yes \_\_\_\_\_

No \_\_\_\_\_

Why? \_\_\_\_\_

\_\_\_\_\_

6. On a scale from 1 to 5, how autonomous have you become?

1. Very dependent
2. Dependent
3. Little autonomous
4. Very autonomous
3. Totally autonomous

## Appendix B. Teacher’s Journal

Observer’s name:	Session:	Date:	Time:	Implementation Stage:	Number of students:
Summary and objectives of the class activities:					
<p><b>What went well?</b>                      Why? How do you know?                      Give evidence from:</p> <ul style="list-style-type: none"> <li>- the planning</li> <li>- the lesson pace</li> <li>- the activities carried out</li> <li>- the students’ production</li> <li>- class atmosphere</li> </ul>	<b>Description</b>				
<p><b>What didn’t go so well?</b>                      Why? How do you know?                      Give evidence from:</p> <ul style="list-style-type: none"> <li>- the planning</li> <li>- the lesson pace</li> <li>- the activities carried out</li> <li>- the students’ production</li> <li>- the class atmosphere</li> </ul>					
<p>What would you do differently, if anything at all, if you were to teach this lesson again?                      Why?</p>					



## Appendix C. Writing Sample

**Write about the following topic:** Overpopulation of urban areas has led to numerous problems. Identify one or two serious ones and suggest ways that governments and individuals can tackle these problems.?

**Give reasons for your answer and include any relevant examples from your own knowledge or experience. Write at least 250 words. (Approx. 4 paragraphs of 5 or 6 lines)**

### Excerpt 17. Student's writing sample

## OVER POPULATION

Like his name says overpopulation is when ^ are a lot of people in the world or in a country and do not exist the necessary implements or the space for all these people. This is a serious problem because it can produce a low quality of life, and the biggest problem is that anybody make a solution for this having a lot of possibilities for decrease (VI) it. Another problem is that teenagers get pregnant in an early age because they don't buy pills for preventing.

The central problem is about teenagers because they dont use pills and don't think in the consequences that this will give you. A solution for this is that schools implement some campaigns for them and explain the consequences, but if we don't regulate it, this could produce a regular or a low quality of life. For example, you can see in traffic lights some poor people asking for some coins and sometimes they have a bigger family but they couldn't give them a good quality of life because children need a lot of things for have (VI) good health conditions. This last one could be one of the consequences of overpopulation if we don't prevent it early. One possible solution for this problem is that the States build some free schools, foundations, hospitals and houses for poor people because they really need it.

Overpopulation requires a lot of implements to guarantee to teenagers a good quality of life for example: it needs better infrastructures for hospitals with an advanced technology and where the State could help or give some facilities for single mothers or housewives that don't have the necessary support for keep (VI) their families. It's incredible that a woman of 15 years old gets pregnant so child and in some cases is because they don't have the required time or attention for they parents. For example one family have (VI) 2 children and the higher (WW) one it' s 15 years old and her sister is 10 years old, their parents work all day and on weekends they only sleep and they don't worry about their kids. So in weeks the higher

one go out with their friends and she comes back before they parents arrive to their house. Months later, she met a guy that she likes so now they are boyfriends. She goes to his house and there passed some acts that they didn't have plan but time later she fell into account realize that they didn't use protection, so now she is mother. She decided to talk with his boyfriend and said what it was the consequence. The guy decided to go out to the country and never call her. She has to accept the consequence and face the situation and get ahead. She new family and their parents in this moment fall into account that their children need more attention.

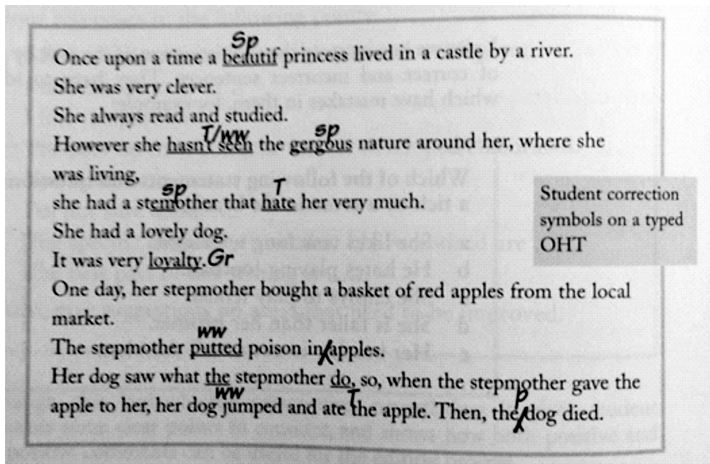
Finally, we have to find a solution for this problem if we dont want to see the terrible consequences that this produce. I believe that overpopulation is a the biggest problem that if we dont react in this moment we could have serious consequences because the world don't have the enough space for all these people, the hospitals dont reach and the States does not have the necessary resources for maintain (VT) all of them.

## Appendix D. Correction Code

1. Look at the following table and read the example errors. Then, look at the next picture to see an example of how the writing assignments are revised.

### CORRECTION SYMBOLS LIST

Symbol	Meaning	Example error
SP	Spelling mistake	The <u>answer</u> is <u>obvius</u> .
VT	Verb tense	I <u>have seen</u> him yesterday.
WO	Word order	I bought a <u>house red</u> .
WW	Wrong word	I am very <u>sensible</u> .
C	Capital letter	I live in <u>miami</u> .
^	Something is missing	He told ^ that he was worry.
Ø	Something is not necessary	He was not <u>more stronger</u> .
?	It is unclear	That is a <u>very excited photograph</u> .
Win a salary	Collocation mistake	Diego <u>win a salary</u> monthly.
Interested on	Preposition mistake	Diana is really <u>interested on</u> tablets.



Student correction  
symbols on a typed  
OHT

Correction code and sample taken and adapted from: Harmer, J. (2004). How to teach writing. Harlow, UK: Pearson Education Limited.

## Appendix E. Consent Letter Participants and Parents



UNIVERSIDAD DE LA SABANA  
COLEGIO HISPANOAMERICANO CONDE ANSÚREZ



Carta de Consentimiento Padres de Familia y Participantes

### INICIO

**Investigación conducida por:** Leonardo Aníbal Alba López

Estudiante III Semestre - Maestría en Didáctica  
del Inglés con Énfasis en Ambientes de Aprendizaje  
Autónomo  
Universidad de la Sabana

**Título del Proyecto:** “Configuring Collocations and Prepositions in Essay Writing through a Corpus-based Strategy”  
“Uso de un *Corpus* (COCA) para mejorar el uso de colocaciones y preposiciones en escritura de ensayos”

Señores

**Padres de familia de participantes**

**Participantes**

**Grado 10B**

Ciudad

Cordial Saludo,

La presente tiene como objeto invitar a los estudiantes e informar a los padres de familia de los posibles participantes en el estudio de investigación titulado “Corpus as a Tool for Improving the Presentation Aspect of Essays in a B1 Group”. Dicha investigación pretende explorar el impacto que tiene el uso de un corpus como herramienta de ayuda para el mejoramiento de las habilidades escritas de los estudiantes, así como determinar cuán autónomos se vuelven los estudiantes cuando se utiliza dicha herramienta para corregir sus trabajos escritos. De ser aprobada la investigación, esta se llevará a cabo durante el cuarto periodo académico del año en curso entre los meses de septiembre a noviembre en el horario de clase habitual.

Durante la investigación los estudiantes recibirán 4 talleres de capacitación relacionados con las temáticas de la investigación. El primer taller tiene que ver con el reforzamiento de conocimientos relacionados con aspectos lingüísticos como lo son las colocaciones y el uso de

preposiciones en inglés. Posteriormente, recibirán un taller relacionado con una metodología llamada “process approach” para desarrollar habilidades escritas. Adicionalmente, los estudiantes recibirán una capacitación integrada relacionada con la escritura de ensayos en inglés y los diferentes usos de un *corpus* para mejorar la calidad de trabajos escritos. Finalmente, los estudiantes empezarán una fase de producción escrita en la cual deberán aplicar los conocimientos adquiridos durante los talleres de capacitación.

Los datos para la investigación se recogerán a través de diarios hechos por el docente, encuestas a los estudiantes y los escritos que los estudiantes hagan durante la fase de producción.

Los estudiantes decidirán participar en el estudio de manera voluntaria a través de una carta de consentimiento. Es importante aclarar que no habrá consecuencias negativas si los estudiantes resuelven no participar o disponen desvincularse de la investigación en cualquier momento. Es decir, los datos obtenidos de él/ella/ellos/ellas no serán utilizados.

De igual manera, cabe aclarar que los estudiantes siguen asistiendo a sus horas de clase de inglés normalmente. Sin embargo, tres de las siete horas que toman semanalmente, se destinarán a todo lo relacionado con la investigación en mención. Asimismo, se les informa que esta investigación cuenta con el total apoyo de las directivas del colegio.

Finalmente, toda la información que los estudiantes suministren será mantenida bajo estricta confidencialidad. Los resultados y conclusiones de la investigación serán presentados únicamente en reuniones profesionales o publicadas en respetadas revistas educativas, pero el nombre y/o cualquier información que pueda identificar a los participantes no serán revelados. Los participantes serán tratados anónimamente.

Si existe preguntas relacionadas con la investigación y su desarrollo, por favor comuníquese con el profesor Leonardo Aníbal Alba López, encargado del proyecto y docente del Colegio Hispanoamericano Conde Ansúrez.

---

Leonardo Aníbal Alba López  
Estudiante – Investigador  
Maestría en Didáctica del Inglés con  
Énfasis en Ambientes de Aprendizaje Autónomo  
Universidad de la Sabana  
Celular: 319 4389190

Correo Personal: lelopez85@gmail.com  
Correo Institucional: leonardo.alba@hispanoamericano.edu.co

## DESPRENDIBLE DE AUTORIZACIÓN

SÍ autorizo \_\_\_\_\_

NO autorizo \_\_\_\_\_

\_\_\_\_\_  
\_\_\_\_\_

Nombre y firma de los padres  
de familia y/o acudientes

Nombre y firma del estudiante