

## Corpus Applications in ELT in Colombia: An Exploratory Survey

Uso de corpus para la enseñanza del inglés en Colombia: una encuesta exploratoria

Rodrigo A. Rodríguez-Fuentes

Valentina Concu

Universidad del Norte, Barranquilla, Colombia

This study examines the application of corpus linguistics in English language teaching (ELT) in Colombia, where its practical adoption remains limited compared to many developed countries. Data were gathered through a survey of English instructors at Colombian universities to assess their familiarity with corpus linguistics. Findings indicate significant challenges, including lack of formal training, low awareness of corpus-based methodologies, and insufficient resources for implementing corpus linguistics applications. The study highlights the potential for enhancing corpus linguistics research and application in Colombia, emphasizing the necessity for improved training and collaborative practices, which could foster greater participation in global corpus linguistics initiatives and enhance the overall effectiveness of ELT.

*Keywords:* corpus linguistics, Colombia, direct corpus applications, indirect corpus applications, English language teaching

Este estudio examina la aplicación de la lingüística de corpus en la enseñanza del inglés en Colombia, donde su adopción práctica sigue siendo limitada en comparación con varios países desarrollados. Se recogieron datos mediante una encuesta a profesores de inglés en universidades colombianas para evaluar su familiaridad con la lingüística de corpus. Los resultados indican importantes retos, como la falta de entrenamiento formal, el poco conocimiento de metodologías basadas en corpus y la carencia de recursos para implementarlas. El estudio resalta el potencial para mejorar la investigación y la aplicación de la lingüística de corpus en Colombia, subrayando la necesidad de una formación adecuada y de prácticas colaborativas, lo cual podría aumentar la participación en iniciativas globales relacionadas con corpus y mejorar la eficacia de la enseñanza del inglés.

*Palabras clave:* lingüística de corpus, Colombia, aplicaciones directas de corpus, aplicaciones indirectas de corpus, enseñanza del inglés

---

Rodrigo A. Rodríguez-Fuentes  <https://orcid.org/0000-0001-7055-0794> • Email: [rodrigof@uninorte.edu.co](mailto:rodrigof@uninorte.edu.co)

Valentina Concu  <https://orcid.org/0000-0003-0863-9602> • Email: [vconcu@uninorte.edu.co](mailto:vconcu@uninorte.edu.co)

How to cite this article (APA, 7th ed.): Rodríguez-Fuentes, R. A., & Concu, V. (2025). Corpus applications in ELT in Colombia: An exploratory survey. *Profile: Issues in Teachers' Professional Development*, 27(1), 115–131. <https://doi.org/10.15446/profile.v27n1.112653>

This article was received on January 25, 2024 and accepted on October 14, 2024.

This is an Open Access article distributed under the terms of the Creative Commons license Attribution-NonCommercial-NoDerivatives 4.0 International License. Consultation is possible at <https://creativecommons.org/licenses/by-nc-nd/4.0/>

## Introduction

Thanks to the advantages of technology, corpus linguistics is becoming a frequent topic of interest in applied linguistics. A meta-analysis by Boulton and Cobb (2017) provided evidence that using corpora is beneficial for learners of foreign languages, particularly regarding effectiveness as measured by effect size in experimental designs. But when confronted with the actual number of language learners (and instructors) of a foreign language, the use of corpora as a means of language learning is still rare or limited, even in some countries with good educational infrastructure and well-established teacher training programs (Breyer, 2009; Ebrahimi & Faghih, 2017). Although corpus linguistics has become increasingly mainstream as a subfield of applied linguistics, the applications of this approach to language studies have not been balanced worldwide in terms of scope, depth, and access (Boulton & Cobb, 2017). Mostly, the state-of-the-art research in corpus linguistics has been carried out by linguists in countries with solid Gross Domestic Product (GDP). In these countries, the stipend for research, in general, can support small but constant research endeavours. In the case of corpus linguistics, these endeavours foster advancements not only for the understanding of pragmatic variations in discourse (Landert et al., 2023) but also for making language patterns (and use) accessible to a broader audience, with direct social implications (Collins et al., 2022; Semino & Short, 2004; Staples, 2015). The advances in corpus linguistics go hand in hand with the access to equipment, technology, and economic resources institutions can offer their members. Most of these institutions are in the Global North, and most members speak English as their L1.

Despite the greater prevalence of corpus linguistics research in the Global North, many countries in the Global South, including those in South America, are implementing direct or indirect corpus linguistics approaches to foreign language teaching, mainly in English as a foreign language (EFL). These countries are

using those corpus linguistics tools originally designed by L1 English users and are adapting them to their classrooms (Al Zumor, 2021; Zhang, 2022). In contrast, due to the relatively short history of accessible foreign language instruction, EFL in South America can still be considered an emerging discipline in many aspects. This history includes recent political efforts from the governments in the region to include English as the main foreign language in the K–12 syllabi. Also, upon graduation, most universities require students to have a functional EFL proficiency level (Hobbs, 2019) that allows them to communicate effectively in different situations or use different basic language functions. Nevertheless, the results are controversial in the best-case scenarios and disappointing in most cases, as the EFL scores in international proficiency measures show (Education First, 2022; Educational Testing Service, 2021). These discouraging results reflect the state of affairs of EFL in South America, where issues beyond the scope of the educational agenda, such as food security, school infrastructure, access to education, and teacher training, among others, limit countries' ability to focus on the best strategies to enhance EFL learning. For instance, the meta-analysis by Boulton and Cobb (2017), which is based on 64 studies, does not include empirical research papers from South America, as there were no articles from this region on the use and effectiveness of corpus linguistics in language learning.

Even so, thanks to globalization, virtual training, and soft data-driven learning materials, there is a sense of vibrant “curiosity” among academic circles to adapt corpus linguistics materials to EFL in South American communities (Sardinha, 2011). There have been academic events<sup>1</sup> to explore applications of corpus linguistics in EFL, English as a medium of instruction (EMI), and content and language integrated learning (CLIL).

<sup>1</sup> For instance, the *III Congreso Internacional de Lingüística Computacional y de Corpus* in 2020 at Universidad de Antioquia (<https://cilcc20.wordpress.com/>) and the Workshop on Automatic Text and Corpus Processing (<http://www.wopatec.cl/>)

These uses and adaptations, however, do not seem to have been documented either as formal interventions or academic articles. Hence, the current investigation explores possible corpus linguistics applications in the EFL Colombian context as an exploratory endeavour to understand the reasons for the lack of participation of South America in the global academic conversation on corpus studies applied to foreign language instruction. More specifically, this study investigates if and to what extent corpora are used in EFL classrooms in Colombia by answering this research question: Are there any direct and/or indirect corpus applications for EFL, CLIL, or EMI in Colombia?

## Literature Review

### Research in Corpus Linguistics in the Colombian Context

As mentioned above, Boulton and Cobb's (2017) meta-analysis included no empirical research papers from South America. After an extensive search in Colombia, several factors might partially explain this lack of papers, including the filters applied by the authors (i.e., not including articles in languages other than English, papers with data collection via observations, papers with unsuitable design for the type of study proposed, and papers that lacked publicly available data to calculate effect sizes). These criteria, as noted by Boulton and Cobb, were implemented to mitigate potential biases, including the inclusion of studies from lesser-known or regional outlets and PhD dissertations, while excluding MA theses. They also acknowledged the limitation of focusing solely on studies written in English, thereby excluding research published in other languages, such as Spanish or Portuguese. More importantly, according to the data collection by Boulton and Cobb, up to mid-2016, there were no published papers from Colombia (or South America) in outlets with a global audience that reported empirical studies (with effect sizes) that used

corpus linguistics or data-driven learning for foreign language learning.

Our survey and extensive search identified a few papers addressing corpus linguistics applications to EFL in Colombia. We found two corpus-based studies published before 2017: One study applied to constructing an EFL test (Trace & Janssen, 2014), and the other reported creating a linguistic corpus to compare speech acts between EFL learners and L1 English speakers (Escobar, 2015). Among the studies published after 2017, we found the study by Pardo Rodríguez (2020), published in Spanish, that described the computer-based search for error patterns in EFL in a learner corpus created by the author. Another study by Nausa (2020) analysed the use of pronouns in a corpus made of 58 transcriptions of oral presentations by Colombian PhD students taking an English for academic purposes course. Although these four studies are empirical, valuable, and published in peer-reviewed journals, they were outside the scope of Boulton and Cobb's (2017) meta-analysis due to their methodological features and lack of effect size reporting. The most recently published study about using corpus linguistics in Colombia (Rodríguez-Fuentes & Swatek, 2022) reported an experimental design that compared the effect size of an indirect approach to corpus linguistics (using corpus-informed textbooks) on a grammatical construction in college EFL learners. These five studies are the only literature in corpus linguistics applied to EFL from Colombia so far.

When expanding the search beyond Colombian borders, we found that in South America, there have been published studies that applied direct and/or indirect approaches in EFL (Arellano, 2018, in Chile; Mussetta & Vartalitis, 2013, in Argentina; Viana, 2006, in Brazil). Survey responses further revealed that corpus linguistics is being applied in Colombia in other fields of applied linguistics and in teaching Spanish as a foreign language.

## Corpus Applications in the EFL Classroom

In the past years, online corpora have become increasingly accessible to a broader audience of scholars and linguists (Biber et al., 1998). Corpus linguistics continues to help explore and redefine theories of language that were difficult to grasp before the creation of corpora. This increased availability also prompted a fruitful convergence between language corpora and teaching (McEnery et al., 2006).

A relevant classification of corpus linguistics for this study is the notion of direct and indirect pedagogical corpus applications to foreign language instruction. Direct corpus linguistics applications are “hands-on” for teachers and learners: They use and interact with actual data and corpora to find patterns. This direct application usually requires access to some type of corpus software to analyse it, as well as knowledge of how to interact with the corpus (Römer, 2011).

Since linguistic corpora started to be used in the second language classroom, they have been seen as an instrument to expose students to a more realistic language use. However, linguistic corpora not only serve as a gateway to a more authentic language but also intend to make students aware of how speakers “really” use the language in different communicative contexts. Several studies have shown that using corpora may also help learners improve their ability to notice and make inferences about new forms and patterns (Gilquin, 2021; Smart, 2012, 2014). They also allow learners to gain first-hand experience working with authentic data on a large scale in a foreign language (Boulton & Cobb, 2017).

Indirect uses of corpora in teaching are primarily seen in the development of material and language testing, alongside the implementation of corpus-informed textbooks and dictionaries in the classroom (McEnery et al., 2006). Due to the data-based nature of corpus-informed materials, the indirect corpus approach has slowly influenced the field of applied linguistics over

the last four decades (McEnery & Xiao, 2010). Despite the use of corpora in the structure and content of reference publishing (e.g., dictionaries) for more than two decades, publishing houses have only (relatively) recently developed textbooks based on corpus-based information (Boulton, 2011). In this sense, textbooks using corpus-based information utilize a student-centred approach and language discovery but have failed to incorporate “an inductive approach to grammar learning” (Smart, 2014, p. 186). Far from being a shortcoming, using a deductive approach signifies inclusion and progress in the field. Corpus-informed materials—though relatively new—have rapidly increased in both number and robustness alongside the growth of corpus linguistics over the last decade. Thanks to international publishing houses, the most widely available corpus-based resources for language instructors are those that present language information for deductive use (e.g., textbooks, dictionaries, worksheets). These convenient materials are particularly beneficial for instructors who lack training in corpus linguistics and have indirectly facilitated the implementation of data-driven learning in language teaching (Rodríguez-Fuentes & Swatek, 2022). In fact, all corpus-based approaches applied to the classroom, direct and/or indirect, have established a dialogue with different language teaching approaches to provide examples of the functional use of language over curricular rules or themes. Boulton (2017) argues that the real effect of data-driven learning on teaching and learning has to do with a new classification of content based on patterns with the eventual emergence of topics, which could be presented in a way that represents current language use.

## Method

### The Survey

We created a survey (see Appendix) to gather information about the knowledge and experiences of EFL teachers and instructors with corpus linguistics in the

Colombian academic context. We sent it via email to 43 active English departments/programs in Colombian universities, asking EFL instructors about their awareness and experience with corpus applications. In each department, EFL program directors, coordinators, and faculty (as listed on their website) were contacted directly. They were also asked to disseminate the survey to any faculty, lecturer, or instructor in their academic division who may have not received it. The survey was built using the branch logic options of Qualtrics (<https://www.qualtrics.com/>), took an average of 15 minutes to complete, could be anonymous, and was divided into four parts. The first part gathered the respondents' general background information: age, education, qualifications, and general experiences in EFL.

The second part of the survey addressed, in general terms, the use of corpus technology in the classroom. Based upon the assumption that linguistic corpora expose students to authentic language and are positively linked to improvements in learners' autonomy, language pattern recognition, and data-driven approaches, the first question in this part aimed to collect information regarding how important corpus-related elements were for EFL teachers in the Colombian context. This part also questioned how comfortable teachers and instructors were with technology and which technological tools they usually implemented in the classroom. The last question addressed the familiarity with corpus linguistics in EFL, EMI, and/or CLIL. If the answer to this question was negative, the respondents were directed to the final part of the survey, in which they were asked if they were interested in taking specialized classes and/or participating in workshops to learn how to implement corpus approaches in their EFL classrooms. Conversely, if the answer to the last question was affirmative, EFL teachers were directed to the third part of the survey.

The third part focused on the applications of corpus linguistic approaches in the classroom. Survey respondents were asked questions to determine if and to what degree they used corpora in the classroom. More

specifically, the questions asked about the language skills (speaking, listening, reading, and writing) and language aspects (grammar, vocabulary, discourse, and translation) for which the respondents considered corpus more useful. For both questions, the respondents could select multiple answers if two or more options were more representative of their experience, and they were also allowed to leave optional comments and list additional aspects not mentioned in the survey.

The following question inquired if the respondents had ever used corpus linguistic approaches in their classrooms. If the answer was negative, the next question asked about the reasons behind this response. Survey respondents could select multiple reasons, such as the lack of training or issues related to time, access, or academic support. In case of a positive answer, the respondents were asked in which classes (EFL, EMI, and/or CLIL) they used corpus approaches. Next, the respondents were asked about the type of corpus approaches (direct and/or indirect) they used in the classrooms. Depending on their answer, EFL teachers were then solicited to indicate which corpora they used (in case they selected "direct corpus approaches") and/or which corpus-based materials they implemented in their classes (in case they selected "indirect corpus approaches"), or whether they had used both approaches.

The last two questions of this third section inquired about corpus linguistics research in Colombia. These questions aimed to gather information about empirical scholarly works that EFL teachers were directly involved with (as authors or co-authors) and whether they knew of other research in corpus linguistics in the Colombian setting. In both cases, EFL teachers were asked to provide a title, link, or DOI.

The survey concluded by asking all the respondents, independently of their answers, if they were interested in taking specialized classes and/or participating in workshops to learn how to implement corpus approaches in their EFL classrooms. The last question solicited

comments, or information EFL teachers thought the authors might need to know.

As an incentive to elicit more responses, survey takers could leave their email addresses to participate in a drawing for one of ten gift cards valued at 100,000 COP (approximately 25 USD).

The survey data were analyzed using quantitative descriptive statistics, mainly percentages of frequency based on the participants' responses.

## Results

### Survey Data

The survey was sent to 311 emails. We received 71 responses from faculty members, lecturers, and instructors across 43 active English departments/programs in Colombian universities. The response rate was 23%, and the data included some of Colombia's most well-regarded English departments.

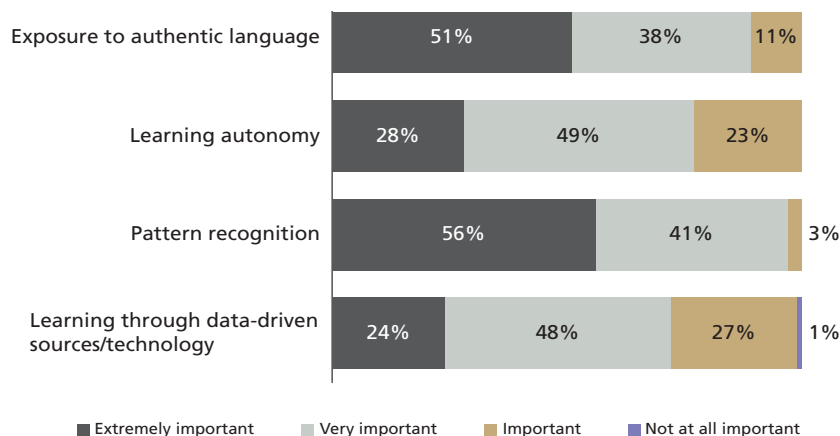
### Background Information

We gathered responses from 37 (52%) EFL instructors who identified themselves as women and 34 (48%) as men. The average age was 40. Of the 71 respondents, 11 (15%) held a BA or an equivalent degree, 48 (65.7%) held an MA, and 12 (16.4%) reported having a doctoral degree. Regarding the question about further qualifications such as TESOL or Corpus Linguistic Training, 24 responded positively. The international teaching certifications reported were TESOL (7 responses), ICELT (3), DELTA (1), English for Specific Purposes (1), and TEFL (1), while others reported having a teaching certificate from their local institution (4). Other respondents reported having an English language certificate such as APTIS for teachers (1), IELTS (1), and Michigan English Test (1), while two of them declared having a certificate in corpus linguistic methodology. On average, the respondents had 11 years of EFL teaching experience.

### Use of Technology in the Classroom

One piece of information we wanted to gather from our respondents before asking them about their familiarity with corpus linguistic approaches was how much importance they gave to those elements that are usually presented in the literature as hallmarks of corpus linguistics: students' exposure to authentic language, improvement in learning autonomy, language patterns recognition, and data-driven learning. Figure 1 shows the responses collected.

Eighty-nine per cent of the respondents believe that exposing learners to authentic language in their classroom is extremely (51%) or very important (38%). These data indicate how aware EFL teachers in Colombia are of the benefits of using authentic language samples to support their teaching. Seventy-seven per cent of the respondents valued in high regard students' learning autonomy (extremely or very important), demonstrating a desire to plan their lessons from a student-centred teaching approach, which has gained a more prominent role in the foreign language classroom since the early 1980s (Crumly et al., 2014; Giannotti 2015; Taylor, 1983). Increasing students' ability to recognize language patterns is also extremely (56%) or very (40%) important among EFL teachers in Colombia, which could also be related to the emphasis on the learners' role in their learning process. Furthermore, there seems to be a need for students to identify lexical bundles, collocations, and other patterns that might support students' learning processes. The importance of learning using technology for the best interest of learners is less widely acknowledged by the instructors, who believe that it is extremely (24%) or very (48%) important. It is worth noticing that this is the only category for which one respondent selected the option "not at all important." The fact that most of the respondents answered "very important" and "important" could suggest that EFL teachers in Colombia are familiar (even inadvertently, in some cases) with the advantages offered by data-driven

**Figure 1.** Instructors' Opinions on the Importance Given to Corpus Linguistics Features ( $N = 71$ )

learning. However, the answer of one participant<sup>2</sup> selecting “not at all important” could imply that there is a (maybe a marginal) part of EFL teachers in Colombia who do not consider data-driven approaches as a valuable tool for the teaching of English.

Another piece of information referred to how comfortable EFL teachers felt with technology in the classroom. The data show that 92% of the respondents feel either comfortable (54.9%) or extremely comfortable (38%) using technology, while 7% declared to feel neither comfortable nor uncomfortable. Only 1% selected extremely uncomfortable as their answer. As mentioned for the question about data-driven learning, this type of answer could indicate that some EFL teachers in Colombia do not use technology-supported tools in their classrooms. This result may be related not only to a lack of interest or training but also to a lack of the necessary infrastructure that makes it possible for EFL teachers to access such technological resources in their educational setting.

In the data, how familiar EFL teachers in Colombia are with technology was further supported by 58

responses (81%) indicating the use of computer-assisted tools in the classroom. The familiarity with technology was also reinforced by optional comments left by the respondents when asked about the tools they used in their class. Fifty-four EFL instructors listed various technological tools, and among the most popular ones are Kahoot! (10 responses), Google Suite (10), YouTube (9), Quizizz (5), Padlet (4), and Mentimeter (4).

### Corpus Linguistics in the Colombian EFL Classrooms

Our main goal was to establish to what extent EFL teachers in Colombia are acquainted with the applications of language corpora in the classroom. Hence, we first posed a yes/no question about the instructors' familiarity with the field of corpus linguistics applied to EFL, EMI, and CLIL. As previously explained, the respondents who answered negatively were directed to the last section of the survey. Our data show that 19% ( $n = 13$ ) of the respondents were unfamiliar with the use of corpus linguistics in EFL, while 81% ( $n = 58$ ) indicated familiarity with this field. To gather more precise information about their knowledge of corpus linguistics approaches in EFL, 81% of respondents who answered positively were then asked two additional questions. The first one inquired about what language skills (writing,

<sup>2</sup> The same participant (who reported having an MA and 20 years of experience in EFL) also answered negatively to both the question about using technology in the classroom and the question about familiarity with corpus linguistic approaches in EFL, EMI, and CLIL

speaking, reading, listening) they considered would be improved through corpus linguistics (the respondents were allowed to select more than one option).

From the 58 respondents, 151 answers were collected. In most cases, the respondents selected more than one skill. According to the data, instructors consider corpus linguistics approaches beneficial for all language skills. Writing (44 responses) and reading (42) were the skills leading the preferences. Listening was the least popular (28), suggesting that many respondents do not see the value of corpus linguistics for this skill. Speaking obtained 37 responses.

The second question asked respondents which language aspects (discourse, translation, vocabulary, and grammar) they considered could benefit from using corpus approaches. For this question, respondents could select more than one option as well, and they also had the opportunity to list other aspects that were not among the choices provided.

From the 58 respondents, 147 answers were collected. The most popular choice was vocabulary (45 responses), followed by discourse (38), grammar (34), and translation (25). Five responses referred to other language aspects. Such answers probably reflect that EFL instructors in Colombia are aware that corpus linguistics might allow them to concentrate on language use in naturally occurring texts instead of focusing on specific structures as abstract concepts separated from how speakers use them. According to Biber et al. (1998), one of the tenets of corpus linguistics is that it provides access to the systematic ways speakers use linguistic features in association with other linguistic and extra-linguistic elements, offering a more accurate and authentic picture of language use. The survey data suggest that EFL instructors in Colombia seem to know what language corpora can offer. This claim is further corroborated by the optional comments left by the respondents: pronunciation, language use, phraseology, writing patterns, genre, rhetorical and stylistic aspects, material design, and comparative linguistics.

All these aspects could be found in most corpora since they often provide extensive collections of texts from different genres and sources (Barth & Schnell, 2021).

To gather insight into the possible existing corpus applications among EFL instructors in Colombia, respondents were then asked if they had ever used corpus linguistic approaches in their EFL, EMI, or CLIL classes. Eighteen respondents (31%) answered negatively. These 18 EFL instructors were then asked about the reasons behind their answers. For this question, there was also the option to select more than one answer among the five items provided. Respondents could also list additional reasons as possible answers to this question. The most popular responses were: “I have no/little training in corpus linguistic approaches” (10 responses), “lack of community of practice to discuss or include corpus approaches” (9), and “lack of technical or academic support in my context” (8). Lack of access or time to plan lessons with corpus approaches were also selected with five and two responses, respectively.

Thirty-three EFL instructors answered that they had used corpus linguistic approaches in their classes. Because of their positive answers, they were presented with a specific set of additional questions. The first one asked about the courses in which they used corpus approaches. Many of them indicated having used corpus approaches in their EFL (27 responses), CLIL (16), and EMI (4) courses. The question that followed asked if these approaches were direct or indirect. Eleven respondents reported using both direct and indirect approaches, 12 only indirect, and 11 only direct.

When respondents indicated using direct approaches, they were asked to choose (from a list provided by the authors) or indicate the corpus or corpus toolbox they used. The data showed that the two most frequently used corpora were COCA (Corpus of Contemporary American English), with 16 responses, and the BNC (British National Corpus), with 11 responses. Five respondents indicated having used their corpus or a corpus administered by their higher education



institution.<sup>3</sup> When respondents stated that they had used indirect approaches, they were asked to indicate which corpus-based material they had used in their classroom. The data showed that the most popular corpus-based materials among EFL instructors are corpus-informed textbooks (15 responses), worksheets from online sources (14), and corpus-informed sources/dictionaries (11).

### Closing Questions

The last two survey questions were shown to all 71 respondents, independently of the answers they gave for the previous questions. The first asked about possible interest in taking specialized classes or workshops about how to implement corpus approaches in EFL, EMI, or CLIL. The response to this question was overwhelmingly positive (69 responses), while only two declared no interest in such an activity. The fact that the vast majority of respondents expressed interest shows how open they are to learning to incorporate direct and indirect corpus linguistics approaches in their classrooms. It is also an indicator that, in the near future, it could be possible to create communities of practice for corpus linguistics in the EFL Colombian context.

The last question was open, asking respondents if there was anything they wanted the researchers to know. We collected nine comments, ranging from the wish for interinstitutional collaborations to suggestions to create workshops for corpus applications in CLIL and general corpus applications in rural areas.

### Discussion

In this section, we will discuss the data obtained through our survey to answer the research question: Are there any direct and/or indirect corpus applications in Colombia in EFL, CLIL, or EMI?

Starting with the formally documented studies reported as research papers produced in Colombia using direct and/or indirect corpus approaches in EFL, CLIL, or EMI, the data from the survey and the independent search by the authors show, in general, a scarcity of empirical research that explains the lack of studies from Colombia in the meta-analysis by Boulton and Cobb (2017). The publication of only five empirical studies, some of them after 2017, shows that research in corpus linguistics applied to foreign languages in Colombia is still at seminal stages and has a long way to go in terms of data, methodological approaches, and massification of knowledge. Colombia, like most South American countries, is currently investing heavily in EFL due to its instrumental value for its inhabitants across social issues (Ministerio de Educación Nacional, n.d.). However, corpus linguistics remains locally unexplored as an approach to improving learning and creating new knowledge.

Based on the survey responses, we found that while most respondents declared to be familiar with corpus linguistics approaches in EFL, EMI, and CLIL, 19% of respondents are not familiar with them. The remaining 81% of the respondents indicated that they believe corpus linguistics approaches benefit all four language skills (especially writing and reading) and different language aspects (vocabulary and discourse, among others). They also reported knowing and using direct or indirect (or both) corpus approaches in the classrooms. Based on this information, the answer to our research question is affirmative; there are corpus applications in EFL, EMI, and CLIL in Colombia. Although most instructors take advantage of the free corpus linguistics tools for EFL instruction for the direct approach—such as COCA and the BNC—a small number of respondents also indicated using their own corpus. This is the case of the publicly available Colombian learners' English Corpus (Pardo Rodríguez, 2020), which implies that corpus linguistics applied to EFL is an emerging field in Colombia.

<sup>3</sup> One of the corpora listed for this response is the CLEC (Colombian Learner English Corpus), developed by María Victoria Pardo Rodríguez and the TNT research group, Escuela de Idiomas, Universidad de Antioquia (<http://grupotnt.udea.edu.co/CLEC>).

On the other hand—and as expected in a context with the previously described limitations—indirect approaches to corpus linguistics (corpus-informed textbooks, worksheets from online sources, and corpus-informed dictionaries) are more popular than direct ones. Higher education institutions in Colombia typically access materials from publishing houses that emphasize strong support from corpus linguistics research, often based on corpora from the US or UK. This means that teachers do not need to invest extra time and effort to learn, train, or obtain the equipment and software required to interact directly with the corpora. The data collected from our survey generally indicate positive attitudes regarding corpus linguistics among EFL instructors in Colombia, who are potential users of this methodology in their classrooms.

The two main reasons listed by those EFL instructors who reported having little or no knowledge of corpus linguistics approaches in EFL were lack of training and lack of a community of practice, issues shared by instructors in other parts of the world (Breyer, 2009; Callies, 2019; Ebrahimi & Faghih, 2017). The limited number of instructors trained in the direct corpus linguistics approach applied to EFL/CLIL/EMI classrooms is further evidenced by our survey, in which only two respondents reported having a certificate in corpus linguistic methodology. This highlights the scarcity of opportunities available in Colombia for the broader EFL instructor community to engage with corpus linguistics. Thus, this creates a fertile ground for higher education institutions worldwide to support horizontally and partner with their Colombian peers to undertake projects that represent the needs of a vibrant community of EFL learners. This idea is also supported by an overwhelming majority (over 97%) of respondents who expressed interest in receiving training in corpus linguistics applied to EFL.

## Conclusions

This study explored possible corpus linguistic applications and research in the context of EFL in

Colombia. The data collected through a survey applied to EFL instructors in higher education institutions demonstrated that both direct and indirect corpus approaches are present in EFL Colombian classrooms yet have limited (for now) uses and scope. In this context, empirical research has just emerged in the last seven years with five studies.

As for the indirect use of corpus linguistics in Colombia, it should be borne in mind that, as Reppen (2010) points out, the difference between corpus-informed and traditionally structured materials is the contents rather than the format of the associated learning activities. Based on this premise, the contents of a corpus-based textbook link the most salient grammatical structures of the corpus to learning activities. The degree of dependence and reliance that instructors might place on textbooks usually makes them take claims (of authentic language samples) of corpus-informed materials by the publishing houses as valid. These claims usually lead to the incorporation of textbooks in the syllabus, assuming that the learners' needs will be met. This is particularly important because, depending on the context, the influence of a textbook might vary from being an extra resource in the class to the foundation of the syllabus and course objectives (Richards, 2012) or deciding the structure and sequence of an EFL course. The potential of corpus linguistics for EFL instruction in developing countries with limited access to educational resources, such as Colombia, presents indirect applications of corpus linguistics as the most immediate and organic way to integrate this approach into EFL.

Direct corpus linguistics applications (O'Keeffe et al., 2007), such as instructors using the tools offered by online corpora to investigate the co-occurrence of specific words or structures, remain largely limited in Colombian EFL classrooms. Cases in which learners were exposed to corpora that contain spoken or written texts that belong to a particular genre (academic papers, private correspondence, literature texts, etc.) are present

but remain mostly unreported in journals with a global audience.

Nonetheless, in South America and particularly in Colombia, there are hopes for the future, as a few respondents reported manuscripts in development, being reviewed, or in-press, that explore corpus linguistics applied to EFL/CLIL/EMI. Likewise, some published studies have reported on the pedagogical exploration of specialized corpora in the EFL context (Arellano, 2018; Escobar, 2015; Mussetta & Vartalitis, 2013; Pardo Rodríguez, 2020; Viana, 2006). Other studies used existing corpora to build items for an EFL proficiency test (Trace & Janssen, 2014) or to assess the effectiveness of corpus-informed vs. non-corpus-informed EFL grammar textbooks (Rodríguez-Fuentes & Swatek, 2022). All these studies highlight the interest in corpus applications in the EFL context in South America. While most of the pedagogical applications of corpus linguistics do not end up as research papers, we have evidence to assert that there are experiences of direct and indirect corpus use in EFL classrooms in this context.

Lastly, it is imperative to address that, based on the responses to the survey, most EFL/CLIL/EMI instructors in Colombia are not formally trained in using corpus linguistics for their courses. Moreover, 19% of respondents are unfamiliar with corpus linguistics as a field of study. In this sense, creating corpus linguistics training and local or transnational communities of practice might bring about more opportunities for support and systematic corpus linguistics interventions. It is only a matter of time before a vibrant community of corpus linguistics users and researchers produce more empirical academic articles that reflect on issues related to EFL learning in Colombia.

### **Further Research**

In most countries with high GDPs, linguists can count on a continuous stream of funding that fosters steady advances in this field. Conversely, scholars who

work in countries with lower GDPs are limited in the depth and scope of the corpus linguistics research they can carry out at their home institutions. This unbalance in the production of empirical research around the world highlights some critical aspects.

An important area of inquiry involves examining strategies prominent international corpus linguistics researchers might use to create a more inclusive academic community. This could include developing strategies to ensure scholars from countries with lower GDPs actively contribute to the corpus linguistics community. It is essential to contemplate what methods and practices can be adopted to democratise access to corpus linguistics research, making it a truly global and collaborative field. One way to encourage such inclusion might begin with the step-by-step integration of corpus linguistics approaches into preservice teacher education (Breyer, 2009; Callies, 2019; Ebrahimi & Faghieh, 2017).

Furthermore, the limitations faced by developing countries, particularly in terms of access to extensive corpora and technology, highlight the need for innovative approaches in corpus linguistics research. It becomes crucial to explore how indirect corpus applications can be incorporated into empirical research projects, especially in foreign language teaching. The aim is to navigate and leverage these limitations to include diverse contexts in the broader corpus linguistics research narrative.

In addition, the challenges associated with limited resources and technological access in certain regions cannot be overlooked. Despite these barriers, investigating ways to effectively implement corpus linguistics in language education is an important area of inquiry. This entails looking at how regions with limited access to technology and resources can still successfully apply indirect corpus approaches in language teaching.

Lastly, the potential for comparative studies between leading nations in corpus linguistics research (or countries with educational infrastructure and resources) and those in different contexts is a promising area for future research. There is evidence that the limitations related to

the use of direct corpus linguistics approaches faced by Colombian instructors in their context are not unique (see Callies, 2019, in Germany and Ebrahimi & Faghih, 2017, in Iran). Such studies could provide valuable insights into how various environments with unique limitations and advantages inform and enhance the effectiveness of different corpus linguistics approaches to language learning. These comparative methodologies might reveal new ways in which corpus linguistics methodologies can be adapted and applied effectively across diverse educational and linguistic landscapes.

Overall, the conversation about the global landscape of corpus linguistics research is not just about acknowledging disparities but also about actively seeking ways to bridge these gaps. It involves rethinking existing paradigms and fostering a more inclusive, innovative, and collaborative approach to corpus linguistics research.

## References

- Al Zumor, A. Q. (2021). Stance in advanced academic writing by Saudi EFL postgraduates: A corpus-based study of critique writing. *Journal of Languages and Language Teaching*, 9(4), 371–384. <https://doi.org/10.33394/jollt.v9i4.4288>
- Arellano, R. (2018). A corpus linguistics application in the analysis of textbooks as national teaching instruments of English as a second language in Chile. *Actualidades Investigativas en Educación*, 18(1). <https://doi.org/10.15517/aie.v18i1.31807>
- Barth, D., & Schnell, S. (2021). *Understanding corpus linguistics*. Routledge. <https://doi.org/10.4324/9780429269035>
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511804489>
- Boulton, A. (2011). Bringing corpora to the masses: Free and easy tools for interdisciplinary language studies. In N. Kübler (Ed.), *Corpora, language, teaching, and resources: From theory to practice* (pp. 69–96). Peter Lang.
- Boulton, A. (2017). Data-driven learning and language pedagogy. In S. Thorne & S. May (Eds.), *Language, education and technology: Encyclopedia of language and education* (pp. 181–192). Springer. [https://doi.org/10.1007/978-3-319-02237-6\\_15](https://doi.org/10.1007/978-3-319-02237-6_15)
- Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning*, 67(2), 348–393. <https://doi.org/10.1111/lang.12224>
- Breyer, Y. (2009). Learning and teaching with corpora: Reflections by student teachers. *Computer Assisted Language Learning*, 22(2), 153–172. <https://doi.org/10.1080/09588220902778328>
- Callies, M. (2019). Integrating corpus literacy into language teacher education: The case of learner corpora. In S. Götz & J. Mukherjee (Eds.), *Learner corpora and language teaching* (pp. 245–264). John Benjamins. <https://doi.org/10.1075/scl.92.12cal>
- Collins, L., Brezina, V., Demjén, Z., Semino, E., & Woods, A. (2022). Corpus linguistics and clinical psychology: Investigating personification in first-person accounts of voice-hearing. *International Journal of Corpus Linguistics*, 28(1), 28–59. <https://doi.org/10.1075/ijcl.21019.col>
- Crumly, C., Dietz, P., & d'Angelo, S. (2014). *Pedagogies for student-centered learning: Online and on-ground*. Augsburg Fortress. <https://doi.org/10.2307/j.ctt9moskc>
- Ebrahimi, A., & Faghih, E. (2017). Integrating corpus linguistics into online language teacher education programs. *ReCALL*, 29(1), 120–135. <https://doi.org/10.1017/S0958344016000070>
- Education First. (2022). *EF English Proficiency Index: A ranking of 111 countries and regions by English skills*. <https://bit.ly/3UoKtMr>
- Educational Testing Service. (2021). *TOEFL iBT test and score data summary 2021*. <https://bit.ly/3ZQvpdT>
- Escobar, W. (2015). Language configurations in the spoken production of Colombian EFL university students. *Colombian Applied Linguistics Journal*, 17(1), 114–129. <https://doi.org/10.14483/udistrital.jour.calj.2015.1.a08>
- Giannotti, J. (2015). *Voices of experience: How teachers manage student-centered ESL classes*. University of Michigan Press. <https://doi.org/10.3998/mpub.7882670>
- Gilquin, G. (2021). Using corpora to foster L2 construction learning: A data-driven learning experiment. *Inter-*

- national Journal of Applied Linguistics*, 31(2), 229–247. <https://doi.org/10.1111/ijal.12317>
- Hobbs, D. (2019, September 26). Functional language and why we should teach it to exam students. *Cambridge ELT*. <https://shorturl.at/x3wTv>
- Landert, D., Dayter, D., Messerli, T. C., & Locher, M. A. (2023). *Corpus pragmatics*. Cambridge University Press. <https://doi.org/10.1017/9781009091107>
- McEnery, T., & Xiao, R. (2010). *Corpus-based contrastive studies of English and Chinese*. Routledge. <https://doi.org/10.4324/9780203847954>
- McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. Taylor & Francis.
- Ministerio de Educación Nacional. (n.d.). *Programa Nacional de Bilingüismo*. <https://shorturl.at/T63NT>
- Mussetta, M., & Vartalitis, A. (2013). Corpus linguistics (CL) in the design of English for academic purposes (EAP) courses. *Frontiers of Language and Teaching*, 4, 45–53.
- Nausa, R. (2020). Identity projection in the oral presentations of Colombian PhD students: Disciplinary differences. *System*, 94, 102351. <https://doi.org/10.1016/j.system.2020.102351>
- O’Keefe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511497650>
- Pardo Rodríguez, M. V. (2020). La construcción de un corpus computarizado de errores escritos con textos de estudiantes universitarios en Colombia. *Lingüística y Literatura*, 41(78), 35–54. <https://doi.org/10.17533/udea.lyl.n78ao2>
- Reppen, R. (2010). *Using corpora in the language classroom*. Cambridge University Press.
- Richards, J. C. (2012, July 4). *The role of textbooks in a language program*. <http://www.professorjackrichards.com/articles/role-of-textbooks/>
- Rodríguez-Fuentes, R. A., & Swatek, A. M. (2022). Exploring the effect of corpus-informed and conventional homework materials on fostering EFL students’ grammatical construction learning. *System*, 104, 102676. <https://doi.org/10.1016/j.system.2021.102676>
- Römer, U. (2011). Corpus research applications in second language teaching. *Annual Review of Applied Linguistics*, 31, 205–225. <https://doi.org/10.1017/S0267190511000055>
- Sardinha, T. B. (2011). Corpus linguistics in South America. In V. Viana, S. Zyngier, & G. Barnbrook (Eds.), *Perspectives on corpus linguistics* (pp. 29–46). John Benjamins. <https://doi.org/10.1075/scl.48.o3ber>
- Semino, E., & Short, M. (2004). *Corpus stylistics: Speech, writing and thought presentation in a corpus of English writing* (1st ed.). Routledge. <https://doi.org/10.4324/9780203494073>
- Smart, J. (2012). *Innovative approaches to ESL grammar instruction* (Publication No. 3524406) [Doctoral dissertation, Northern Arizona University]. ProQuest Dissertations and Theses.
- Smart, J. (2014). The role of guided induction in paper-based data-driven learning. *ReCALL*, 26(2), 184–201. <https://doi.org/10.1017/S0958344014000081>
- Staples, S. (2015). *The discourse of nurse-patient interactions: Contrasting the communicative styles of U.S. and international nurses*. John Benjamins. <https://doi.org/10.1075/scl.72>
- Taylor, B. P. (1983). Teaching ESL: Incorporating a communicative, student-centered component. *TESOL Quarterly*, 17(1), 69–88. <https://doi.org/10.2307/3586425>
- Trace, J. W., & Janssen, G. (2014). Corpus-informed test development: Making it about more than word frequency. *Shiken: JALT Testing and Evaluation SIG Newsletter*, 18(1), 3–9.
- Viana, V. (2006). Modals in Brazilian advanced EFL learners’ compositions: A corpus-based investigation. *Profile: Issues in Teachers’ Professional Development*, 7(1), 77–86.
- Zhang, J. (2022). The “Chinese dream” as a discursive symbol: A corpus-assisted discourse study on political discursive construction and global media representations of the “Chinese dream”. *Chinese Semiotic Studies*, 18(4), 605–631. <https://doi.org/10.1515/css-2022-2082>

### **About the Authors**

**Rodrigo A. Rodríguez-Fuentes** holds a PhD in Second Language Studies from Purdue University (USA). He is an Assistant Professor in the Department of Foreign Languages at Universidad del Norte, Colombia. His research interests include the relationship between technology and language learning, as well as classroom assessment and testing procedures applied to EFL.

**Valentina Concu** holds a PhD in Linguistics from Purdue University (USA) and is an Assistant Professor in the Department of Foreign Languages at Universidad del Norte, Colombia. Her research interests include historical corpus linguistics, historical pragmatics, complexity theory, complex network science, and second language acquisition.

## Appendix: Survey on Corpus Linguistics in Colombian EFL College Classrooms

1. Age
2. Gender
3. Academic degree
4. Years of experience teaching English
5. Do you have further qualifications relevant to ELT (e.g., TESOL Certification, Corpus Linguistics training, or other types of certifications)?
  - Yes (specify)
  - No
6. According to your experience, how important (*Not at all important, Important, Very important, Extremely important*) is it to improve EFL students'
  - exposure to authentic language?
  - autonomy?
  - language pattern recognition?
  - learning through data-driven sources/technology?
7. How comfortable are you with the use of technology in the classroom?
  - Extremely uncomfortable
  - Uncomfortable
  - Neither comfortable nor uncomfortable
  - Comfortable
  - Extremely comfortable
8. Do you use computer-assisted tools in your EFL classroom?
  - Yes
  - No
9. Are you familiar with the field of corpus linguistics applied to English as a foreign language (EFL), English as a medium of instruction (EMI), or content and language integrated learning (CLIL)?
  - Yes
  - No
10. For which language skill(s) do you consider corpus approaches valuable? (Select all that apply)
  - Listening
  - Speaking
  - Reading
  - Writing
11. For which language aspect(s) do you consider corpus approaches valuable? (Select all that apply)
  - Grammar
  - Vocabulary
  - Discourse
  - Translation
  - Other (specify)

12. Have you ever published empirical or theoretical research (journal articles, book chapters, conference proceedings, thesis, or dissertations) using the corpus linguistics approach(es) applied to EFL, EMI, or CLIL with data collected in Colombia?
  - Yes. Please share the name of the author(s), piece of research, link, or DOI.
  - No
13. Do you know of any professor who has published empirical or theoretical research (journal articles, book chapters, conference proceedings, thesis, or dissertations) using the corpus linguistics approach(es) applied to EFL, EMI, or CLIL with data collected in Colombia?
  - Yes. Please share the name of the author(s), piece of research, link, or DOI.
  - No
14. Have you ever used corpus linguistics approaches in your (EFL, EMI, or CLIL) teaching practice?
  - Yes
  - No
15. Why do you not use corpus linguistics approaches in your (EFL, EMI, or CLIL) teaching? (Select all that apply)
  - I have no (or little) training in corpus linguistic approaches.
  - I do not see the value of corpus linguistics in EFL, EMI, or CLIL.
  - Lack of time to plan lessons with corpus approaches.
  - Lack of access to direct or indirect corpora.
  - Lack of technical or academic support in my context.
  - Lack of community of practice to discuss or include corpus approaches.
  - Other (specify)
16. Which corpus linguistic approach have you used in your (EFL, EMI, or CLIL) classes? (Select all that apply)
  - Direct (where students are encouraged to use and explore the corpus tools themselves)
  - Indirect (where materials are based on corpus findings and neither students nor teachers need to interact with the corpus)
17. Which corpora have you used? (Select all that apply)
  - COCA (Corpus of Contemporary American English)
  - British National Corpus
  - LancBox (Lancaster University Corpus Toolbox)
  - CROW (Corpus & Repository Writing)
  - Local Learner Corpus
  - My own (or my university's) corpus data. Please share the name, link, or brief description.
  - Other limited and/or freely available corpus online. Please share the name, link, or brief description.
18. Which type of corpus-based material have you used? (Select all that apply)
  - Worksheets from online sources
  - Corpus-informed textbooks (e.g., *Touchstone*, *Valid Choice*, *Face2Face*, *Grammar and Beyond*, *English in Mind*, *Interactive*, *English Unlimited*, *Viewpoint*, *Unlock*, or others).
  - Corpus-informed dictionaries



- Frequency graph, concordance lines, lexical bundles, etc., from corpus-informed sources
  - Other (specify)
19. I use corpus-based approaches in my (select all that apply):
- EFL courses
  - EMI courses
  - CLIL courses
20. Would you like to learn in specialized classes/workshops how to implement corpus-based approaches in your EFL, EMI, or CLIL classroom?
- Yes
  - No
21. Is there anything else you would like the researchers to know?
- Yes (specify)
  - No