



Data-Driven Learning in an EFL class: a study of Ecuadorian learners' perceptions

Aprendizaje basado en datos en la enseñanza de inglés como idioma extranjero: percepciones de estudiantes ecuatorianos

Hover Ismael Conza-Armijos

Universidad Nacional de Loja, Loja, Ecuador

hover.conza@unl.edu.ec

 <https://orcid.org/0000-0001-6182-0991>

Liliana Fernanda Celi-Celi

Universidad Técnica Particular de Loja, Loja, Ecuador

lfceli@utpl.edu.ec

 <https://orcid.org/0000-0003-1681-2380>

Recepción: 17/06/2023 | Aceptación: 29/08/2023 | Publicación: 10/09/2023

Cómo citar (APA, séptima edición):

Conza-Armijos, H. I., y Celi-Celi, L. F. (2023). Data-Driven Learning in an EFL class: a study of Ecuadorian learners' perceptions. *INNOVA Research Journal*, 8(3), 37-50.
<https://doi.org/10.33890/innova.v8.n3.2023.2297>

Abstract

Previous research has reported an increased interest in the application of Data-Driven Learning (DDL), which was derived from corpus linguistics, to aid English as a foreign or second language learning. DDL emerged as a learner-centered approach that emphasizes the discovery of linguistic patterns by analyzing natural occurrences of language. Many studies have investigated the effect of DDL on several linguistic components and have reported positive outcomes. Consequently, its application has germane importance in an Ecuadorian context as it has not been widely explored. This study seeks to document the experience of 14 EFL students from an Ecuadorian public university and examine their perceptions towards the implementation of DDL activities. The intervention included two stages: assisted guidance for students to retrieve information on their own and corpus-based materials developed by the teacher. Students' perceptions were gathered via a questionnaire, a reflection, and a follow-up interview. The findings indicate that students

perceived DDL as a practical and innovative approach to support grammar discovery in an EFL class. Nevertheless, students encountered some challenges associated with the use of DDL, for example, the limited number of searches, an issue that can be resolved by a complete registration in the COCA web. This article concludes with pedagogical implications and recommendations for future DDL applications and research in Ecuadorian EFL classes.

Keywords: COCA; corpus linguistics; data-driven learning (DDL); EFL class; learners' perceptions.

Resumen

Muchas investigaciones han divulgado la implementación del aprendizaje basado en datos, conocido en inglés como Data-Driven Learning (DDL), que surgió de la lingüística de corpus para contribuir con el aprendizaje del inglés. El DDL se enfoca en el estudiante y fomenta el descubrimiento de patrones lingüísticos a través de un análisis de ejemplos suscitados en un contexto natural. Numerosos estudios han investigado los efectos de DDL en ciertos componentes lingüísticos y han reportado resultados positivos. Por lo tanto, la implementación de DDL es crucial en un contexto ecuatoriano ya que hasta el momento no se ha explorado. Este trabajo documenta la experiencia de 14 estudiantes en una universidad pública ecuatoriana y recoge sus percepciones sobre la implementación de actividades con el DDL. Esta intervención se realizó en dos etapas: la primera, incluyó una capacitación acerca del uso de corpus para obtener ejemplos de forma independiente. La segunda etapa estuvo a cargo del docente y consistió en la creación de materiales basados en corpus. Con respecto a las opiniones de los estudiantes se utilizó un cuestionario, una reflexión y una entrevista. Los resultados mostraron que los estudiantes consideran el DDL como un enfoque práctico e innovador para coadyuvar con el aprendizaje de la gramática del inglés. Algunas limitaciones giraron en torno al número restringido de búsquedas que la herramienta permite, lo cual, se resolvería con un registro previo en la web de COCA. Este artículo concluye con recomendaciones pedagógicas para aplicaciones e investigaciones futuras de DDL en un contexto similar.

Palabras claves: aprendizaje basado en datos; COCA; inglés como lengua extranjera; lingüística de corpus; percepciones de estudiantes.

Background to the study

Corpus linguistics

Lindquist (2009) defines corpus as "a collection of texts which is stored on some kind of digital medium used by linguists to retrieve linguistics items for research" (p 3). For Crawford and Csomay (2016), corpus stores a large number of texts (written or spoken) from where the study of language in use can be conducted (see also Boulton, 2012; O'Keefe, McCarthy, and Carter, 2007). In this realm, the study and analysis of these data is conceptualized as corpus linguistics (Cheng, 2012).

Over the past few years, the contributions of corpus linguistics have spread to many fields, one of which is language teaching because it provides complete insights into how people use the language (Conrad, 2010). Huang (2011) asserts that corpus linguistics enables distinguishing language forms at various levels of formality. By the same token, this comprehensive analysis shows how different speech functions help language users across contexts. This is supported by

Biber and Reppen (2002) who claim that individuals do not often use language as prescribed in grammar books and that lexical meanings are contingent on contexts and users.

Currently, there are different types of corpus that are representative of certain aspects of language that can be employed to analyze, describe and teach the language. For example, the British National Corpus comprises more than 100 million words that represent the universe of contemporary British English. The Corpus of Contemporary American English (COCA) (Davis, 2008) is the largest of its kind and contains more than one billion words of text.

COCA

The source of data for this work is COCA (<https://www.english-corpora.org/coca/>) largely because learners in this context are taught American English, so it is a good way of exposing them to more authentic input. COCA is well equipped with analytical tools that enrich the search such as frequency, charts for genres and years, collocates, comparisons, KWIC, POS tags, etc. In the pedagogical world, these tools allow manipulating the language, which provokes learning and retention (Reppen, 2010).

One of the main tools of COCA is a word concordance called 'Key-Word-In-Context' (KWIC). KWIC enables viewing corpus data in which samples of real language are presented in lines of text known as concordances. Users can also access an extended context for each occurrence of the input search (Pérez-Paredes et al., 2019). This tool helps identify aspects of a particular word or phrase such as lexico-grammatical patterns or semantic prosody. (Boontam & Phoocharoensil, 2018).

When retrieving results with KWIC, the word occurrences are displayed vertically with the node word in the center of the line and the surrounding context highlighted in different colors. The colors signal the part of speech as follows: orange for adverbs, purple for verbs, turquoise for nouns, yellow for prepositions, green for adjectives, blue for pronouns, and grey for articles, demonstratives, and connectors. Learners may find the structural analysis more manageable provided that the interface is visually appealing.

Data-driven learning in ELT

Corpus Linguistics has made evident that language exhibits a highly patterned structure. A direct application of this language characteristic in teaching is the Data-Driven Learning (DDL) approach, which involves examining patterns in a target language and inferring meaning and use (Johns, 1991). In this seminal work, Johns captures three central moments towards language analysis: identify, classify, and generalize. Identifying refers to the recognition of the target structures. It then leads to a systematic classification of the retrieved information, which the teacher or the student can do. Finally, generalizing entails a profound analysis to draw logical conclusions regarding language rules.

DDL offers greater insights to be applied in foreign language environments because it compiles natural and authentic examples of language in use (Reppen, 2010; Luo, 2016). Within DDL, learners play a pivotal role as they become "language detectives" (Johns, 1997, as cited in

Flowerdew, 2012) because they need to recognize and examine the emergent patterns to draw their conclusions (Smart, 2014).

In addition, DDL favors a guided inductive approach to language learning where the elements of explanation and corpus use are suited to students' needs (Johansson, 2009; O'Keeffe, 2020). In this respect, the teachers' roles entail guiding students whilst students perform the analysis (Boontam & Phoocharoensil, 2018). There are two possible methods to exploit DDL; for one, teachers can prepare the materials in advance or direct students to conduct the searches independently. Consequently, the learners' role is twofold: they act as language learners and as language researchers simultaneously (Cheng, 2010).

Guan (2013) and Smirnova (2017) assert that DDL fosters problem-solving skills as learners explore and discover language structures, unlike teacher-centered methodologies. Szudarski (2018) highlights the role of corpus in increasing learning autonomy and understanding real-life linguistic communication. This is because corpus does not interpret data; it is rather the duty of the researcher (Szudarski, 2018; Gavioli, 2005). The outcomes derived from this process result in a richer and more meaningful language learning experience.

Previous Studies

A great deal of research has investigated the application of DDL in EFL settings and students' attitudes towards the use of corpus-based learning methodology. For example, in 2014, Geluso and Yamaguchi conducted a study on DDL use to discover formulaic language for a 15-week period. This study took place in a private university in Japan and involved 29 students at the A2-B2 CEFR level. For data collection, the authors employed a questionnaire, follow-up interviews, and student reflection logs. The findings revealed that students had positive attitudes regarding DDL use despite some minor difficulties such as unfamiliar vocabulary.

A study by Boontam and Phoocharoensil (2018) examined the effectiveness of paper-based DDL method for the learning of English prepositions with 35 A2 students at a private school. The participants were taught the prepositions using DDL for about 6 weeks. The researchers gathered data through a pre and post-test and complemented the process with a questionnaire and an interview. The results depicted those students found the learning experience with DDL fun, interesting, and challenging.

A small-scale study by Yoon and Hirvela (2004) reported positive findings in the implementation of a corpus approach in an American university. These 22 international students viewed the corpus approach as valuable to support the L2 writing process. In addition, participants pointed out that DDL boosted their confidence to develop this productive skill.

A recent longitudinal study by Lin and Lee (2019) analyzed the pedagogical use of DDL for EFL grammar classes in Taiwan. Three instructional formats were used to compare students' grammar performance: a traditional deductive approach (TDA) only, a combination of 40% DDL with 60 % of TDA, and a blend made of 60% DDL plus 40% of TDA. It was found that all the treatment arrangements helped students to learn grammar. However, in terms of attitudes, students tended to favor the DDL approach.

Overall, these previous studies show that DDL may produce a meaningful learning experience in EFL contexts. To this date and to the best of our knowledge, no studies have been undertaken in Ecuadorian institutions to implement DDL and collect students' or teachers' perceptions. For this reason, we deem necessary to conduct a study to capture students' perceptions towards DDL and, by the same token, to shed light on pedagogical applications.

Research Questions

This study attempts to answer two research questions:

- What are Ecuadorian EFL university learners' perceptions of adopting a DDL approach for grammar learning?
- What approach to learning grammar do students prefer?

Methodology

Participants

The participants in this study were taking general English lessons as a requirement prior to obtaining their formal qualification in an Ecuadorian public university. Students would normally attend face-to-face sessions but given the health emergency caused by the Covid-19 pandemic, they received virtual lessons via Zoom. The deliberate instruction took place twice a week for 60 minutes per session, where grammar teaching was paired with language skills development.

This small-scale research took place in an elementary English class, Level 4 from the EFL program. The class consisted of 23 students, of whom 14 students volunteered to participate (11 female and 3 male). The participants were majoring in teaching and education, and their ages ranged from 20 to 24 years old. All participants shared the same nationality and L1 (Spanish). They had received English for 6 years as part of Ecuador's national curriculum for secondary education and nearly 300 hours of instruction in the English language program.

DDL Treatment and materials

At the beginning of the semester, students took a diagnostic test that was an adaptation from the Key A2 by Cambridge Assessment English (2020). The test originally consisted of reading, writing, listening, and speaking, but the speaking part was removed due to practicality issues. This test served the purpose of gauging students' language proficiency.

The treatment was implemented in weeks 5, 6, and 7 of the 16-week semester, and each treatment lasted for two hours. The topic unit was food, and the grammar revolved around countable and uncountable nouns and quantifiers. See figure below:

Figure 1

The screenshot of COCA interface

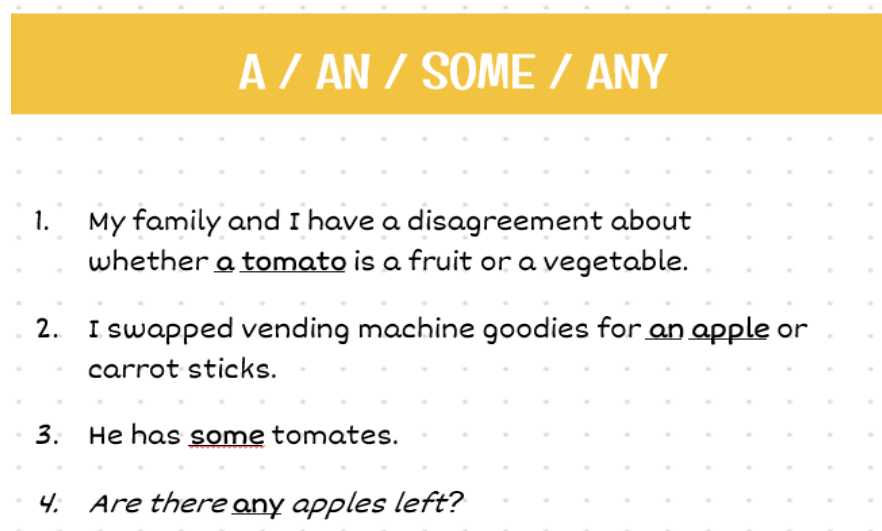


The first synchronous session included the presentation of food semantic sets and activities to promote repeated encounters with the vocabulary. For the asynchronous work, students searched for two examples for the quantifiers a, an, some, and any followed by a food item.

The teacher started the second synchronous session by having students share one of their examples on Google Jamboard. The rationale behind this activity was to guide students' attention to singular and plural nouns before presenting the grammar topic of countable and uncountable nouns. After some guided practice, students were prompted to work in groups of four in Zoom Breakout Rooms. It is worth mentioning that students chose their peers they wanted to work with to create a quiz with 5 multiple-choice questions on Quizziz. For the asynchronous work, students had to take the quizzes created by their peers to allow further practice and consolidate their grasp of the content. Also, to ensure accountability, students had to keep a record of the examples they retrieved on a digital foldable template. The teacher provided this template and included an organized layout and detailed instructions of the tasks at hand.

Figure 2

Students' examples of their COCA search for quantifiers

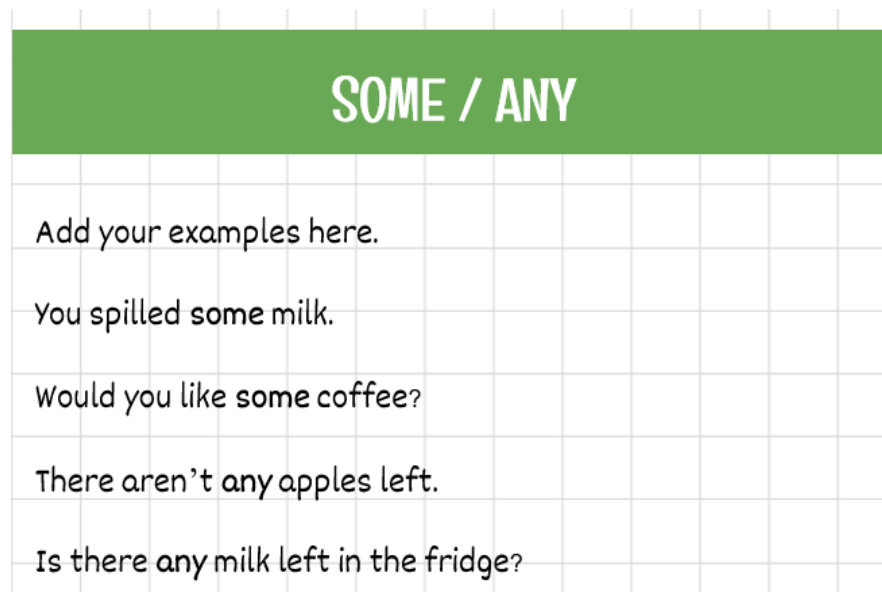


A / AN / SOME / ANY

1. My family and I have a disagreement about whether a tomato is a fruit or a vegetable.
2. I swapped vending machine goodies for an apple or carrot sticks.
3. He has some tomatoes.
4. Are there any apples left?

Figure 3

Students' examples of their COCA search for Some and Any



SOME / ANY

Add your examples here.

You spilled some milk.

Would you like some coffee?

There aren't any apples left.

Is there any milk left in the fridge?

The third synchronous session entailed a lexical review as a lead-in to food containers. Then, to facilitate the content of *How much...?* and the quantifiers *a lot*, *a little*, *not much*, and *none*, the teacher provided students with a handout. It contained contextualized examples taken from COCA along with two key questions to trigger inferences about its use. Once again, students worked collaboratively in Zoom Breakout Rooms to thoroughly analyze and identify patterns. After responding to the questions, the participants had to produce two examples following the

identified patterns. Finally, to check comprehension, the teacher called on some students to share their examples on the Zoom whiteboard and gave corrective feedback when required. As part of the asynchronous activities, learners focused on completing the fourth flap of their digital foldable writing two examples per category.

Figure 4

Example of DDL materials introducing How much

Examples: How much..?

- But did you see **how much sugar**'s in there?
- **How much water** is that?
- **How much milk** should I store at a time?
- **How much meat** do you really eat for breakfast?
- Dad, **how much coffee** have you had?
- **How much salt** did you use? Just a pinch

Are the food items in the examples countable or uncountable?

Based on your analysis, we can infer that how much is used to ask about...

Figure 5

Example of DDL materials introducing quantifiers

Examples: A lot, a little, not much, none

- The rest of my diet is made of poultry, fish, **a little meat**.
- Add **a little butter** or oil and saute the onion, celery, and carrot
- Lunch is fruit - yes, with good bread and sometimes **a little cheese**

- Wow! That's **a lot of sugar**
- Caitlin drinks **a lot of coffee**.
- I don't drink **a lot of beer**.

- We don't use too **much sugar**
- Don't worry I didn't use too **much milk**,
- No, don't drink so **much juice**.

- I just put my dish in the **oven** and I didn't use **any salt**,
- I didn't get **any cheese** on our burgers because you tell me it's no good for me

The last synchronous session complemented the grammar topic of quantifiers, and it followed the same approach as the previous session. Students had to work in Zoom Breakout

Rooms for a deep analysis of examples, but on this occasion, with an emphasis on *How many...?* and the quantifiers *a lot, a few, many, and any*. To wrap up the unit, the teacher reviewed the grammar content and utilized a formative assessment on Quizziz.

Data Collection

To gather data about students' attitudes towards DDL use, we relied on three instruments: a questionnaire, a reflection that was included in a digital foldable, and a follow-up interview at the end of the treatment.

The construction of the questionnaire required a sound review of the literature which led to the selection of 15 statements adapted from Yoon and Hirvela (2004), Geluso and Yamaguchi (2014), and Boontam and Phoocharoensil (2018). The questionnaire comprised four categories: tool, grammar, experience with DDL, and future application. The participants rated the statements using a four-point Likert scale (4: strongly agree, 3: agree, 2: disagree, 1: strongly disagree).

At the end of the intervention, participants had to complete a reflection section mainly to evaluate DDL. The reflection prompts were written in students' L1 so they could express their ideas clearly irrespective of their English proficiency level (A2). This instrument served two purposes: to underpin the creation of the interview questions, and to bridge the information from the other instruments.

Finally, the researchers conducted an individual semi-structured interview with 5 volunteers (4 females and 1 male) to have a more comprehensive view of their experience. The questions were devised to elaborate on the information provided by participants in the questionnaire and the reflections. The researchers interviewed participants in Spanish to aim for a more comfortable interaction and avoid any misunderstandings due to participants' English proficiency level. The recordings were transcribed and then translated into English by the researchers.

Results and discussion

Table 1

Perceptions about the corpus tool

Statement	Strongly agree		Agree		Disagree		Strongly disagree		Mean	Mode	SD
	n	%	n	%	n	%	n	%			
	1. The searching tool KWIC was easy to use.	3	21	10	71	1	7				
2. I had difficulty in retrieving the examples with the target forms.	2	14	8	57	4	29			2,86	3	0.64

The items for the first category are related to the evaluation of COCA, particularly the use of KWIC. The results reflect an overall positive trend regarding the use of KWIC. Further data in the journals showed that the tool is described as "easy to use and intuitive", "user-friendly and manageable", and "free of charge".

However, a great majority mentioned that they had difficulty in retrieving examples, which can be attributed to the little or no previous experience with COCA. These perceptions are supported by the interviewees who stated that they had not used COCA before besides the initial guidance. It was found that, for some students, the interface was complex to navigate at first, they lacked knowledge on how to navigate the tool, and they had to follow multiple steps to conduct the word search. Moreover, they also reported that attempting more than 10 searches constituted a hindrance to retrieving fast information.

Table 2

Perceptions about grammar learning

Statement	Strongly agree		Agree		Disagree		Strongly disagree		Mean	Mode	SD
	n	%	n	%	n	%	n	%			
3. The real texts in the corpus are too difficult to understand.	5	36	3	21	3	21	3	21	2,71	4	1.16
4. The use of COCA was good for understanding the differences between countable and uncountable nouns	7	50	7	50					3,50	4,3 *	0.5
5. I like to discover the language patterns by myself.	5	36	8	57	1	7			3,29	3	0.59
6. Learning through DDL makes me understand the lesson better.	3	21	11	79					3,21	3	0.45
7. I completely understand how to use how much and how many by learning through DDL method.	5	36	9	64					3,36	3	0.48

*Note** Responses occurred equally.

The statements for the second category revolved around grammar learning aided by COCA. It is observed that the responses for the first item about real texts being difficult are evenly distributed among the four options. Yet, more respondents are inclined to the agreement end of the continuum. Despite these results, some participants ascertained that COCA retrieves natural and complex sentences, contextualized and high-quality examples.

Regarding the preferred grammar learning approach (See item 5), the majority of students pointed out that they like to discover language patterns, which favors the inductive approach. To this end, the interviewees claimed that COCA offers well-structured examples, which eases comprehension. Overall, most participants were satisfied with DDL implementation because it enhanced the lesson delivery, which increased understanding of the target grammar.

Table 3*Experience using DDL*

Statement	Strongly agree		Agree		Disagree		Strongly disagree		Mean	Mode	SD
	n	%	n	%	n	%	n	%			
8. I understand the purpose of using the corpus in this course.	6	43	7	50	1	7			3,36	3	0.61
9. The guidance from the teacher was efficient to explore COCA.	9	64	5	36					3,64	4	0.48
10. I think that DDL method is more challenging than other traditional methods	2	14	8	57	3	21	1	7	2,79	3	0.77
11. I was encouraged to direct my own learning, actively think, and express my reflections during the DDL sessions	4	29	10	71					3,29	3	0.45
12. When I learn grammar, I prefer a learner-centred method like DDL.	3	21	7	50	4	29			2,93	3	0.70

This third category sought information pertaining to participants' experience with DDL. One of the most salient details from this section was that participants were aware of the rationale behind the use of corpus. To put it simply, students highlighted that COCA has a plethora of information paired with authentic examples taken from different contexts. Interestingly, students identified that COCA allows for increasing vocabulary as an additional benefit.

Likewise, the majority of participants acknowledged the guidance from the teacher as efficient to explore COCA. It is worth mentioning that students only received one entire introductory session to navigate COCA.

The remaining three items (10, 12, and 13) illustrate that students perceive DDL as more challenging than traditional methods and helpful for fostering autonomous learning. That is the case of one of the interviewees who mentioned that she was encouraged to think critically when choosing examples. Another interviewee believed that DDL stimulated active participation. This is also supported by one participant who wrote in his journal: "DDL promotes independent research".

Table 4

Future application

Statement	Strongly agree		Agree		Disagree		Strongly disagree		Mean	Mode	SD
	n	%	n	%	n	%	n	%			
13. I will use the corpus for my English learning in the future.	2	14	11	79	1	7			3,07	3	0.46
14. I recommend using DDL for the same unit with future cohorts.	5	36	9	64					3,36	3	0.48
15. I want to study other English lessons through DDL method in future units.	4	29	9	64	1	7			3,21	3	0.56

As for the last category about future applications, the majority of the participants indicated that they are likely to employ COCA in the future. In fact, the journal analysis revealed that 12 respondents would use COCA in the long term, and all the interviewees corroborated this. Similarly, most participants agreed with item 14, thus recommending DDL to cover the same unit. Finally, when participants were asked about implementing DDL in future units, the overall response to this question was very positive.

Conclusions and pedagogical implications

This study set out to explore EFL learners' perceptions towards the implementation of a DDL approach for grammar learning in an Ecuadorian university. Based on the information obtained from the questionnaire, journals, and interviews, it is evident that students had positive impressions of the corpus-based approach. The participants described COCA as user-friendly and easy to navigate, and those findings are consonant with the results of Tribble's study in 2015. Besides, participants highlighted the use of corpus as an additional language exposure to a wide range of natural and authentic examples. Nonetheless, some students reported difficulties navigating COCA due to their little or no previous experience with the tool. A common shortcoming was that searches were limited to 10 if users had not registered previously.

Another emerging point derived from this study shows that students prefer an inductive approach because they mentioned that it guides grammar discovery and enhances their comprehension. These results echo the findings of Flowerdew (2015) who believes that DDL and a discovery-based approach are inexorable given that learners induce language rules from the corpus. Flowerdew further asserts that inductive learning is closely linked with discovery learning and triggers higher-order thinking skills such as noticing, analyzing, and making inferences.

The present study provides additional evidence with respect to the potential benefits of DDL in language learning that researchers have already reported in numerous studies. Learners recognized the usefulness of DDL as it encouraged active thinking about the target grammar. Despite the positive feedback, it is important to note that teachers' time can be compromised when

developing materials and becoming familiar with the corpus tools. We concur with Römer (2007), and we advocate for a DDL-friendly environment for both teachers and students where they can receive appropriate guidance to access corpora. Only then will they be able to exploit other resources that corpus consultation tools offer. Ecuadorian EFL teachers could implement the DDL methodology in an endeavor to promote learner autonomy, particularly useful amidst remote instruction.

Limitations and further research

This study was based on pioneering research and was adapted to an unprecedented fully virtual environment. However, one of the major limitations was the low number of participants; therefore, it is difficult to generalize the findings. Although the intervention was conducted throughout a unit, it was short in length as it only lasted 3 weeks. Further research is called for to explore DDL use with a longitudinal study over a term or a semester. To this extent, a pre and a post-test could be valuable tools to gather quantitative data to support and contrast students' claims.

Provided that this study was solely focused on gathering students' perceptions of DDL use for grammar learning, future research could investigate the effect of DDL on other language components. For example, given the importance of vocabulary for accurate and appropriate language use, studies on collocations could be conducted to increase students' lexical competence. Within the same category, researchers may consider comparing the learning of synonyms and near-synonyms such as make/do or win/gain/earn.

Bibliographic References

- Biber, D., & Reppen, R. (2002). What does frequency have to do with grammar teaching? *Studies in Second Language Acquisition*, 24(2), 199–208. <https://doi.org/10.1017/S0272263102002048>
- Boontam, P., & Phoocharoensil, S. (2018). Effectiveness of English preposition learning through data-driven learning (DDL). *3L: Language, Linguistics, Literature*, 24(3), 125–141. <https://doi.org/10.17576/3L-2018-2403-10>
- Boulton, A. (2012). What data for data-driven learning? *EuroCALL Review*, 20(1), 23–27. <https://bit.ly/484oO1c>
- Cambridge Assessment English. (2021). A2 Key Preparation. <https://www.cambridgeenglish.org/exams-and-tests/key/preparation/>
- Cheng, W. (2010). What can a corpus tell us about language learning. In *The Routledge handbook of corpus linguistics* (pp. 319–332). Abingdon, Oxon: Routledge.
- Cheng, W. (2012). *Exploring corpus linguistics: Language in action*. Banbury: Routledge.
- Conrad, S. (2010). What can a corpus tell us about grammar? In *The Routledge handbook of corpus linguistics* (pp. 227–240). <https://doi.org/10.4324/9780203856949.ch17>
- Crawford, W., & Csomay, E. (2016). *Doing corpus linguistics*. New York: Routledge.
- Davis, M. (2008). The corpus of contemporary American English (COCA). <https://www.english-corpora.org/coca/>
- Flowerdew, L. (2012). Applying corpus linguistics in teaching arenas. *Corpora and Language Education*, 190–232. https://doi.org/10.1057/9780230355569_7

- Flowerdew, L. (2015). *Data-driven learning and language learning theories*. Amsterdam: John Benjamins.
- Gavioli, L. (2005). *Exploring corpora for ESP learning*. Amsterdam: John Benjamins.
- Geluso, J., & Yamaguchi, A. (2014). Discovering formulaic language through data-driven learning: Student attitudes and efficacy. *ReCALL*, 26(2), 225–242. <https://doi.org/10.1017/S0958344014000044>
- Guan, X. (2013). A Study on the application of Data-driven learning in vocabulary teaching and learning in China's EFL Class. *Journal of Language Teaching and Research*, 4(1), 105–112. <https://doi.org/10.4304/jltr.4.1.105-112>
- Huang, L. S. (2011). Corpus-aided language learning. *ELT Journal*, 65(4), 481–484. <https://doi.org/10.1093/elt/ccr031>
- Johansson, S. (2009). Some thoughts on corpora and second language acquisitions. *Corpora and Language Teaching* (pp. 33–44). <https://doi.org/10.1075/scl.33.05joh>
- Johns, T. (1991). Should you be persuaded two samples of data-driven learning. *ELR Journal*, 4, 1–16. <https://bit.ly/45vsGGP>
- Lin, M. H., & Lee, J. Y. (2019). Pedagogical suitability of data-driven learning in EFL grammar classes: A case study of Taiwanese students. *Language Teaching Research*, 23(5), 541–561. <https://doi.org/10.1177/1362168817740899>
- Lindquist, H. (2009). *Corpus linguistics and the description of English*. Edinburgh: Edinburgh University Press.
- Luo, Q. (2016). The effects of data-driven learning activities on EFL learners' writing development. *SpringerPlus*, 5, 1255. <https://doi.org/10.1186/s40064-016-2935-5>
- O'Keeffe, A. (2020). *Data-driven learning- A call for a broader research gaze*. *Language Teaching*, 1–14. <https://doi.org/10.1017/S0261444820000245>
- O'Keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge: Cambridge University Press.
- Pérez-Paredes, P., Ordoñana Guillamón, C., Van de Vyver, J., Meurice, A., Aguado Jiménez, P., Conole, G., & Sánchez Hernández, P. (2019). Mobile data-driven language learning: Affordances and learners' perception. *System*, 84, 145–159. <https://doi.org/10.1016/j.system.2019.06.009>
- Reppen, R. (2010). *Using corpora in the language classroom*. Cambridge: Cambridge University Press.
- Römer, U. (2007). Corpora and language teaching. 112–131.
- 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>
- Smirnova, E. A. (2017). Using Corpora in EFL Classrooms: The Case Study of IELTS Preparation. *RELC Journal*, 48(3), 302–310. <https://doi.org/10.1177/0033688216684280>
- Szudarski, P. (2018). *Corpus linguistics for vocabulary: A guide for research*. London: Routledge.
- Tribble, C. (2015). Teaching and language corpora. In *Multiple affordances of language corpora for Data-driven Learning*. Amsterdam: John Benjamins.
- 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>