

## **Hands-on Practices on the Use of Corpora in English Language Teaching: Reflections from Teacher Training**

**Gamze Çalışkan\***

**Safiye İpek Kuru Gönen\***

---

### **ARTICLE INFO**

Received 17.05.2023  
Revised form 05.10.2023  
Accepted 07.10.2023  
Doi: 10.31464/jlere.1298644

#### **Keywords:**

*Corpus-based language teaching  
and learning  
Teacher views on corpus use  
Training on corpus pedagogy*

---

### **ABSTRACT**

Along with recent advances in the development of corpus tools, there has been a growing interest in integrating corpora into language teaching. However, teachers' lack of awareness may be an obstacle to the effective implementation of corpora in language classroom routines. This study offers an in-service training program on corpus use in language pedagogy and aims to investigate teachers' views on corpus-based teaching and learning regarding their actual hands-on experiences. For this purpose, 10 EFL instructors at a state university in Turkey were enrolled in an intensive seven-week training program. This program involved a total of nine sessions aimed at presenting how to use corpus tools in language teaching and implementing corpus-based activities in vocabulary teaching, grammar teaching, and error correction. Data were collected via reflective logs and semi-structured interviews. Qualitative analyses revealed the importance of providing teachers with in-service training on corpus use since they perceived to gain awareness, knowledge, and skills required for implementing pedagogical applications of corpora in language teaching while also reporting drawbacks of corpus use in practice. The findings of the study shed light on future practices related to pedagogical uses of corpora in language classrooms and offer a number of implications for teacher education.

---

### **Acknowledgments**

This paper is derived from an unpublished M.A. thesis dissertation of the first author under the supervision of the second author as the thesis advisor.

### **Statement of Publication Ethics**

The ethical approval for this research was obtained from Anadolu University Social and Humanities Scientific Research and Publication Ethics Board with the document dated 29/04/2019, numbered 29928.

### **Authors' Contribution Rate**

Both authors contributed equally to the manuscript.

### **Conflict of Interest**

The authors declare that there is no conflict of interest.  
The authors declare that there is no conflict of interest.

### **Reference**

Çalışkan, G., & Kuru Gönen, S.İ. (2023). Hands-on practices on the use of corpora in English language teaching: Reflections from teacher training. *Journal of Language Education and Research, 9(2)*, 317-344.

---

\* Instructor, ORCID ID: <https://orcid.org/0000-0001-6739-046X>, Anadolu University, School of Foreign Languages, [gkarasahin@anadolu.edu.tr](mailto:gkarasahin@anadolu.edu.tr)  
Assoc. Prof. Dr., ORCID ID: <https://orcid.org/0000-0002-7374-8269>, Anadolu University, Department of English Language Teaching, [ipekkuru@anadolu.edu.tr](mailto:ipekkuru@anadolu.edu.tr)

## Introduction

Recent advances in technology and the accessibility of various digital tools have directed a great deal of attention to the idea of corpus use in language pedagogy (Boulton & Cobb, 2017; Huang, 2017). For language teaching and learning purposes, corpora can be implemented indirectly or directly (Römer, 2011, p. 207). Indirect applications refer to the influence of findings stemming from corpus analysis on teaching syllabuses and instructional materials (e.g., textbooks and dictionaries) whereas direct applications involve teachers' and/or students' hands-on interaction with corpora. Teachers can engage students in corpus-induced activities including concordance lines (a line of text taken from a corpus) chosen in advance in the form of paper-based activities (Boulton, 2010), or students have access to corpus data via computers to search for target items and analyze concordance lines (Godwin-Jones, 2017). Such applications that encourage students to discover language patterns by analyzing corpus data are also known as data-driven learning (DDL henceforth) (Johns, 1991).

Research on DDL has mainly focused on studies investigating the effectiveness of this approach in learning vocabulary (e.g., Chen, 2017; Daskalovska, 2015; Karras, 2016; Li, 2017), grammar (e.g., Lin & Lee, 2019; Smart, 2014), and error correction and feedback process in writing (e.g., Crosthwaite, 2017; Luo & Liao, 2015; Luo, 2016; Mueller & Jacobsen, 2016; Quinn, 2015; Tono et al., 2014). These studies highlight that implementing corpora in the areas of vocabulary teaching, grammar teaching, and error correction would benefit learners in gaining awareness of the use of language in context and correction of language errors via authentic samples. In addition to experimental studies, there are several meta-analyses examining the impacts of DDL on learning language skills/areas (Boulton & Cobb, 2017; Lee et al., 2019; Mizumoto & Chujo, 2015). This bulk of research in general reported numerous gains in the learning of language skills and areas such as increased retention of vocabulary, gains in grammatical knowledge, successful error correction, and awareness in the development of language skills.

Apart from research that has focused on the effects of DDL on students' gains in language learning, teachers' attitudes toward using corpus in language learning and teaching environments, as well as training pre-service teachers on corpus use, have been the focus of recent literature (e.g., Abdel Latif, 2021; Ebrahimi & Faghil, 2017; Farr, 2008; Leńko-Szymańska, 2017; Zareva, 2017). As for in-service teacher training, several studies explored teachers' perceptions of corpus use in language pedagogy based on workshops designed (e.g., Chen et al., 2019; Mukherjee, 2004) whereas few studies have aimed to investigate teachers' perspectives on the integration of corpora into language teaching contexts with their actual experiences in classrooms (Lin, 2019; Poole, 2022). The present study offers a comprehensive in-service training program that empowers EFL teachers with corpus knowledge and provides opportunities to produce corpus-based materials and activities to be used in real classrooms. In this way, this study may illuminate the suitability of corpora for actual classroom practices from in-service teachers' perspectives.

## Literature Review

### Use of Corpora and Language Teacher Education

Using corpora in the form of DDL activities is associated with the principles of discovery learning since learners are expected to explore patterns related to language use and infer rules themselves (Boulton & Cobb, 2017; Flowerdew, 2015). Such a discovery process might also be regarded as part of a constructivist approach (Flowerdew, 2015) as learners build knowledge by drawing conclusions about the uses of the language. In addition, corpus applications in language pedagogy align with the noticing hypothesis, proposed by Schmidt (1990) since corpus-based activities direct students' attention to recurring language patterns (Flowerdew, 2015), and a huge amount of authentic data makes language patterns noticeable (Boulton & Cobb, 2017). This approach enhances learners' awareness of the language (Breyer, 2009; Farr, 2008) and might also be motivating since learners are involved in the learning process (Gilquin & Granger, 2010; Lee, 2011). Besides, analyzing data accessed via corpora and identifying language patterns encourages students to take part in the learning process actively fostering learner autonomy (Boulton & Cobb, 2017; Braun, 2005; Godwin-Jones, 2017) as well.

Although there has been a growing enthusiasm for pedagogical uses of corpora in the research area, corpus applications in actual classroom practices are not prevalent (Leńko-Szymańska & Boulton, 2015). Researchers draw attention to teachers' lack of awareness and knowledge about corpora (Boulton, 2009; Gilquin & Granger, 2010; Frankenberg-Garcia, 2012; Römer, 2011). It is essential to empower teachers with basic skills such as making corpus queries and interpreting corpus data to enhance language teaching (Frankenberg-Garcia, 2012). Other than considering corpus-induced pedagogy merely as a remedy or one-shot practice, teachers need sufficient knowledge and necessary skills related to classroom uses of corpus applications to integrate them into actual teaching routines (Leńko-Szymańska, 2014).

It can be seen that integrating corpora into language teacher education is crucial (Farr & O'Keeffe, 2019). In a study, Farr (2008) found that student teachers' views about using corpus as a learning tool were generally positive as they could see the language in its natural context along with some challenges such as the amount of time required and technical difficulties. In the context of an initial language teacher education program in Germany, Breyer (2009) highlighted the positive impacts of the training on student teachers since the participants could see the potential of corpora for language education although several problems related to technology and classroom management were noted. Similarly, Heather and Helt (2012) stressed the significance of training on the basics of the corpus in language teacher education despite some concerns related to technological issues and students' language proficiency levels in actual classrooms. Leńko-Szymańska (2014) conducted a study with graduate students in a Polish context and investigated their perceptions of corpus use in language education. The results stressed the potential benefits of corpus use for vocabulary in general and phraseology in particular. Besides, the emphasis placed on training teachers to be able to gain sufficient knowledge and skills required for corpus use was notable. In a similar vein, Zareva (2017) carried out a study to explore TESOL trainee teachers' perceptions on the integration of corpus-based research in an English grammar course and concluded that taking such a course about

corpus-based research broadened participants' perspectives in many ways although technical challenges were echoed with an emphasis on the need for more hands-on experience. In another study, Ebrahimi and Faghih (2017) examined pre-service teachers' reflections on a seven-week online course at an M.A. program. Along with the need for training on pedagogical applications of corpora, several limitations in relation to corpus use were noted as it required technological equipment, time for designing corpus-based materials, and analysis of concordances. In a recent study, Abdel Latif (2021) concluded that student teachers held positive attitudes toward corpus use after receiving instruction on corpus literacy; however, such positive beliefs did not deeply influence their practices in the long term.

As for the perceptions of in-service teachers, Mukherjee (2004) surveyed 248 language teachers in a German EFL context on the principles of corpus and pedagogical applications of corpora in language classrooms. The results indicated that teachers held positive attitudes toward corpus consultation since it might be useful while preparing teaching materials and checking exam papers. In the Taiwanese EFL context, Lin and Lee (2015) investigated six teachers' perceptions regarding the use of corpus-based activities in teaching grammar and concluded that such activities promoted students' involvement in the lesson since they became active participants. However, concerns regarding workload and the material preparation process were observed. In another study, Lin (2019) explored one teacher's experiences with corpus-based activities in grammar instruction and emphasized the need for more practice to internalize such an approach to language teaching. In a more recent study, Poole (2022) concluded that corpus-based activities were beneficial; however, using ready-made materials was not favored since it did not foster sufficient autonomy to make corpus queries. Concerning this finding, the study showed that it was important to use corpus-based materials designed specifically for the context to maximize the potential benefit of the corpus in language teaching. Besides, it was pointed out that making corpus searches and interpreting findings might not be appropriate for some students because of their lack of proficiency.

In the Turkish EFL context, Özbay and Kayaoğlu (2015) unveiled language instructors' perceptions of using corpus tools in language teaching regarding the training they received. Analysis of qualitative data showed that the training raised teachers' awareness of the accessibility of corpora and the information that could be accessed via corpus tools. Based on the participants' reflections, corpora would help to reach various authentic samples of the language, and they might be useful in checking accurate uses of the language. Albeit potential benefits such as gaining awareness on the availability of corpora and classroom applications of corpora in language education, several challenges such as technical issues, difficulties in material preparation, and students' lack of language proficiency were also notable in the previous research (e.g., Ebrahimi & Faghih, 2017; Heather & Helt, 2012). Thus, the current study adds to the relevant literature on providing training opportunities to teachers to raise their awareness by highlighting a need to guide teachers in material preparation tailored to the needs of their students and teaching contexts.

## **Significance and Aim of the Study**

Despite the increasing popularity of corpus studies in recent years, pedagogical uses of corpora are still not prevalent in Turkish EFL classrooms as teachers lack awareness, knowledge, and skills (Aşık, 2015). Since courses related to corpora are not systematically integrated into language teacher education programs, the familiarity with corpus use is mainly limited to the individual efforts of language educators. Instructors might encounter the term “corpus” as part of their professional development activities while pursuing postgraduate degrees and attending workshops, conferences, and certificate programs; however, using corpora for pedagogical purposes requires training to acquire the essential knowledge and necessary technical skills (Lin, 2019; Römer, 2011). Such training programs need to create opportunities to observe the ways to bring corpora to language classrooms and experiment with classroom applications of the corpus. Very little research has provided systematic training on developing hands-on classroom applications and explored in-service teachers’ perceptions regarding corpus use for pedagogical purposes (e.g., Lin, 2019; Poole, 2022). This study provided an extensive training program and investigated language teachers’ views on the use of corpora in real-life classroom contexts based on their hands-on experiences throughout the training. Therefore, this study addressed the following research question:

R.Q. What do EFL university instructors think about using corpora in their actual teaching contexts upon receiving intensive training on corpus use in language pedagogy?

## **Methodology**

### **Participants and Context**

The ethical approval for the current study was obtained from the ethics board of the institution in which the research was carried out. Before the study, all participants were informed about the procedures of this research and their right to withdraw from this study at any time. To ensure the confidentiality of the data they provided and their voluntary participation, they signed consent forms. Participants were coded to maintain their anonymity (e.g., *T1*, *T2*, *T3*).

The study was conducted at the School of Foreign Languages at a state university in Turkey in the Fall Term of the 2019-2020 Academic Year. In the context of this study, intensive language courses are offered to students from various majors. Based on the results of the placement test administered at the beginning of the fall term, students are assigned to classrooms at four different proficiency levels ranging from beginner to intermediate that are determined according to the Global Scale of English (GSE). Instructional materials used in the preparatory program consist of a published textbook series and supplementary packs compiled by members of the material development unit at the school. These materials primarily used in classrooms do not include any samples of corpus-based activities.

10 EFL university instructors who were working in the study context participated in the study. Participants were selected using convenience sampling (Creswell, 2012) as they were available and willing to receive training on corpus use in language teaching. This study followed a convenience sampling strategy since voluntary participation of the instructors was necessary to spare time for the training and integration of corpus-based language pedagogy.

Among seven participants who were graduates of English Language Teaching (ELT), three of them held an M.A. degree in ELT as well. Three participants had an M.A. degree in non-ELT programs (i.e., English Linguistics). The instructors' years of teaching experience ranged from 5 to 22, and the language proficiency level they were teaching varied from *beginner* to *intermediate*. The background information about all participants can be found in Table 1.

**Table 1.** Background Information About the Participants

	Gender	Education	Experience	Teaching Level
<b>T1</b>	F	B.A. in ELT	22	Beginner
<b>T2</b>	F	B.A. in ELT	20	Intermediate
<b>T3</b>	F	M.A. in ELT	20	Pre-intermediate
<b>T4</b>	F	M.A. in ELT	15	Pre-intermediate
<b>T5</b>	F	B.A. in non-ELT	13	Pre-intermediate
<b>T6</b>	F	M.A. in ELT	9	Pre-intermediate
<b>T7</b>	F	B.A. in non-ELT	7	Beginner
<b>T8</b>	M	B.A. in ELT	7	Beginner
<b>T9</b>	F	B.A. in non-ELT	7	Elementary
<b>T10</b>	F	B.A. in ELT	5	Beginner

To gain an insight into participants' background knowledge and prior experiences with corpora, one-on-one interviews were held with each participant before receiving training (see Appendix A). About the familiarity with corpora, four participants (*T1*, *T2*, *T7*, and *T8*) had almost no idea what a corpus referred to. Three participants (*T3*, *T5*, and *T10*) described the concept of the corpus as a collection of words; however, they were not familiar with any corpora. Three participants (*T4*, *T6*, and *T9*) had experiences with corpora such as the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA) to check the frequency of words and usage of collocations and registers. One of the participants (*T4*) introduced corpora to several students once to show how they could consult corpora to observe various usages of words and check collocations. The initial interviews aided in obtaining information about the participants' background regarding familiarity with corpora and pedagogical uses of the corpus. The findings showed that most of the participants' familiarity with corpora was quite low, and none of them had actual experiences with corpora as part of their regular teaching practices.

### Training Procedures

An extensive training program on the use of corpora in language teaching was developed based on the teachers' need to get acquainted with corpus tools and their pedagogical uses. Specific attention was paid to the cyclical nature of training to create opportunities for participants to observe the information newly presented and experiment with it. That is, participants were first engaged in tasks in the position of learners to become familiar with the new information. Then, they applied the knowledge and skills newly acquired to their actual teaching context. Participants reflected on every single experience in this cyclical pattern and all steps were enhanced with observations, reflections, group discussions, and hands-on experiences. This training was designed in the form of two-hour weekly workshops. The program consisted of nine sessions with specific themes, and the training process was

implemented in seven weeks. All training sessions included interactive tasks, group discussions, and assignments that were guided with thought-provoking questions throughout the training (see Appendix B for sample materials). The themes covered in training sessions each week can be seen in Table 2.

**Table 2.** The Weekly Schedule of Training Sessions

Week 1	Session 1	Overview on corpus
Week 2	Session 2	Introduction to basic functions of the corpus tool: COCA
Week 3	Session 3	How to use concordance lines in classrooms: Lexical items
Week 4	Session 4	Sharing participants' corpus-based activities on lexical items
Week 4	Session 5	How to use concordance lines in classrooms: Grammatical structures
Week 5	Session 6	Sharing corpus-based activities on grammatical structures
Week 5	Session 7	How to use concordance lines in classrooms: Error correction
Week 6	Session 8	Sharing corpus-based activities on error correction
Week 7	Session 9	Discussion on the experience of implementing corpus-based activities

*Session 1: Introduction of basic notions of the corpus.* Various engaging tasks were designed (e.g., fundamentals of DDL, listing words according to frequency, identifying chunks used in written and spoken genres) to attract instructors' attention to the characteristics of corpus and its applications in language teaching and learning. The session was elaborated with sample pages from a corpus-informed textbook to provide insights into how corpus data could be integrated into teaching resources.

*Session 2: Introduction of the Corpus of Contemporary American English (COCA).* This session was carried out in a computer laboratory to empower participants with practical skills. COCA was selected as the tool in the study since it provided a convenient interface, up-to-date information, and free access to the website (Timmis, 2015). Participants acquired hands-on experience in using the tool by accomplishing several exercises (e.g., searching a word/phrase, analyzing concordance lines, comparing and sorting words according to sections) and were familiarized with the interface through continuous support and guidance.

*Session 3: Use of concordance lines in vocabulary instruction.* Following group discussions on vocabulary teaching practices (i.e., strategies, problems, activity types, and the potential benefit of corpus use), participants were provided with various samples of corpus-based activities in vocabulary instruction (e.g., gap-filling, matching, and selecting) for different language proficiency levels. Sample activities were composed of both direct use of raw corpus data via computers and indirect use including printout corpus data that consisted of concordance lines from COCA. At the end of this session, participants were introduced to the task in which they were required to prepare a corpus-based material for the following week.

*Session 4: Reflection on the process of material preparation for vocabulary instruction.* Participants were free to choose any lexical item/s at any proficiency level and they uploaded their materials into a digital shared folder to compose a corpus-based material pool. Participants examined all materials designed by their colleagues and exchanged ideas on corpus-based activities produced for vocabulary instruction. Reflections were carried out on the material preparation process based on their hands-on experiences.

*Session 5: Use of corpus-based materials for grammatical structures.* Participants held a group discussion on their classroom routines while teaching grammatical structures (i.e.,

strategies, activity types, and problems encountered in teaching grammar items) and the potential use of corpus-based materials in grammar instruction. Upon the completion of group discussions, participants were engaged in sample corpus-based activities derived from concordance lines on COCA for grammar instruction. Grammar structures selected for tasks were specified based on teaching syllabuses in the study context (e.g., functions of the modal *can*, passive voice, *since/for* with present perfect tense, and indirect questions). Participants were also asked to design a corpus-based activity for grammar instruction as an assignment for the next session.

*Session 6: Reflection on the process of material preparation for grammar instruction.* Participants followed similar procedures to upload and share corpus-based activities prepared for grammar instruction. Participants reviewed all materials produced by their colleagues and reflected on the material preparation process including the reasons for activity type, target structure, and level they chose and used in their corpus-based activity.

*Session 7: Use of concordance lines in error correction.* Following group discussions on participants' actual practices regarding error correction in writing, this session focused on error correction via corpus tools with direct and indirect uses (e.g., accurate use of structures/prepositions, spelling, and linkers), and the errors selected for tasks were derived from students' writing papers in the context of the study. This practice was followed by the presentation of the following assignment that required the preparation of a corpus-based activity for error correction.

*Session 8: Reflection on the process of material preparation for error correction.* This session was reserved for discussions on the process of corpus-based material preparation to correct students' errors in writing. Adopting similar procedures conducted in sessions 4 and 6, participants examined corpus-based materials they shared and reflected on the use of corpus-based materials for error correction. Participants were also assigned a final task in which they were required to implement one corpus-based activity they prepared for their students. For this task, participants were expected to select one of the areas covered in the training (i.e., vocabulary instruction, grammar instruction, or error correction). In this way, participants had the opportunity to use corpus-based activities in their current classroom environment.

*Session 9: Overall discussion and reflection on the use of corpus-based activities.* Participants reflected on the practices of using corpus in actual teaching contexts. All participants evaluated the applicability of corpus-based activities to the level they were teaching in terms of practicality, effectiveness, gains, and challenges.

## Data Collection

**Reflective logs:** Participants wrote logs following each training session and reflected on what they thought and felt about the session they attended. In this way, instructors' thoughts and feelings about corpus use in language teaching could be elicited based on hands-on experiences they acquired in training sessions. Each participant kept a total of nine digital logs and the researcher created a separate folder for all participants considering their confidentiality.

**Semi-structured interviews:** The interview protocol with each participant was carried out twice: before the training and after the training. Data collected through interviews before the training were descriptive and were used to obtain information about the participants' background and their familiarity with corpora. Thus, interviews before the training were not used to answer the research question regarding the participants' views on using corpora in their teaching contexts. To explore participants' views about the use of corpora in language teaching after receiving training, semi-structured interviews were conducted as they provided flexibility to obtain qualitative data (Nunan, 1992). Two experts in the field of ELT were consulted to ensure the appropriateness of interview questions for the current study (see Appendix C). Upon completion of the training program, interviews were scheduled with each participant for the following week. All interviews were carried out in Turkish since participants could express themselves comfortably in their native language. The interviews were audio-recorded to be transcribed verbatim.

### **Data Analysis**

Qualitative data collected through reflective logs and semi-structured interviews were analyzed using the Constant Comparative Method (CCM) which is based on the ideals of grounded theory for data analysis (Corbin & Strauss, 2015). Using procedures of this method allowed for a theory emanating from data rather than following categories identified beforehand (Kolb, 2012). Data were analyzed in a three-step procedure. In the first step, data were segmented into small pieces to be labeled as a code; that is, words or phrases involving any sense or thought about the perceptions of participants were identified (Allan, 2003). In the second step, the codes were compared with another one constantly in many cycles to find out recurring themes. In the third step, using an ongoing analysis process, assembled themes constituted the final categories of main themes. Two independent raters were involved in the process to ensure the reliability of qualitative data analysis. To overcome any disagreements between the raters and finalize the codes, a negotiation session was carried out. The percentage of agreement between the raters was measured by using the percentage agreement formula (Huberman & Miles, 2002), and it was found .90 indicating a high level of reliability.

## **Results**

### **EFL University Instructors' Views on Using Corpora in Language Teaching upon Receiving Training**

Qualitative analyses of reflective logs and semi-structured interviews revealed a total of 713 codes related to the instructors' views about using corpora in language pedagogy and their overall perceptions of the training. These codes were assembled under three main categories and 16 sub-categories. The main categories were determined as *teachers' perceived gains* (290 codes), *benefits of using corpora in language teaching* (212 codes), and *drawbacks of using corpora in language teaching* (211 codes). In the following sections, tables related to each main category and sub-category are given along with sample explanatory statements. These explanatory statements are participants' expressions extracted from the actual data anonymously to exemplify what each sub-category involves.

### Teachers' Perceived Gains

Findings revealed that instructors in the study perceived numerous positive gains related to their teaching. The main category of *teachers' perceived gains* from training on corpus use was composed of five sub-categories. The sub-categories regarding teachers' perceived gains during the process of training are displayed in Table 3.

**Table 3.** Distribution of Codes Related to Teachers' Perceived Gains

Sub-categories	Explanatory statements	N*
Awareness of corpora	"Tasks encouraged learning different aspects of the corpus."	73
Familiarity with classroom applications of corpora	"I learned corpus is a useful tool to teach English in several other ways."	72
Technical skills required for corpus use	"I learned what to do with COCA and how to use its different functions."	70
Collaboration with colleagues	"It was valuable to learn about my colleagues' experiences and perspectives."	39
Hands-on experience in material preparation	"I had a chance to use COCA for the first time to prepare a task for my students."	36
<b>Total</b>		<b>290</b>

N\*= Number of codes

Background survey before training on corpus use put forward that the instructors were not very familiar with using corpora in language teaching. With the help of the comprehensive training process both on the theoretical and practical uses of corpora, the participants reported that they gained *awareness of corpora* (73 codes) in different aspects, such as the concept of the corpus, the availability and accessibility of corpus tools, and the influence of corpus on language pedagogy (e.g., corpus-informed textbooks, syllabus design, and DDL). Moreover, two participants, who were familiar with corpus, reported that this training broadened their perspectives regarding their knowledge of corpus as they were introduced to the background information and potential of corpus knowledge in language use. One of the prominent gains perceived in this study was gaining *familiarity with classroom applications of corpora* (72 codes). Incorporating corpora into grammar instruction and writing for error correction was a completely new idea for the majority even though several participants had insights into the use of corpora in vocabulary instruction. The participants also stressed that context-specific corpus-based activities were inspiring as they had an in-depth understanding of how to apply corpus use to their actual classroom practices in various ways. The participants acknowledged that training sessions aided them in gaining *technical skills required for corpus use* (70 codes); that is, they were empowered with essential skills required for using corpus comfortably and effectively (e.g., searching words/phrases on the corpus, and analyzing concordance lines, using various functions of COCA). The following excerpts illuminate the perceived gains participants acquired from training sessions:

T9: "Before this training, I had some doubts about the possibility of using corpus in classrooms. I thought that corpus could be used only in teaching collocations. I would have never imagined using corpus in error correction and grammar... I

*particularly liked this training as it involved classroom applications of the corpus.”*

*T3: “Without your help, it would have been very difficult for me to discover the essential tips. I’d tried to use a corpus tool before, but I wasn’t able to figure out how to do it and I never tried it again. So, the tips provided in the sessions were really useful for us.”*

The training sessions also contributed to the participants’ *collaboration with colleagues* (39 codes). Participating in this training increased cooperation among the instructors throughout the study since they shared their thoughts, feelings, concerns, and suggestions about corpus use in language pedagogy through interactive discussions at every step of training. Furthermore, it was noted that having access to corpus-based materials prepared by all participants was a substantial gain as the instructors might use context-specific teaching activities in their future practices. The following excerpt shows how this training fostered collaboration among the participants in this study:

*T8: “Thanks to my colleagues, I learned a lot of things that I wouldn’t have discovered myself. In this sense, it was fun. In each session, I felt like I was back in my undergraduate years. It was nice to communicate with others and discuss our ideas in groups...”*

As T8 reported, collaboration among peers promoted their learning and contributed to their professional lives. In addition to the aforementioned gains, the participants acquired *hands-on experience in material preparation* (36 codes). The instructors reported that it was a valuable experience to get involved with the material preparation process despite the challenges they encountered. Preparing corpus-based activities enabled them to practice the knowledge and skills acquired during the training.

### Benefits of Using Corpora in Language Teaching

The second main category included six sub-categories related to the benefits of using corpora in language teaching from the instructors’ perspectives. The sub-categories regarding the benefits of using corpora in language teaching are displayed in Table 4.

**Table 4.** Distribution of Codes Related to the Benefits of Using Corpora in Language Learning

Sub-categories	Explanatory statements	N*
Effectiveness in error correction	<i>“Students can see and analyze both their own mistakes and good forms of problematic structures.”</i>	69
Usefulness in vocabulary instruction	<i>“Corpus can be useful in teaching confusing words.”</i>	53
Helpful in accessing authentic/reference source	<i>“Corpus can be a good reference to find original sentences while preparing materials.”</i>	53
Usefulness in teaching particular structures	<i>“Corpus can be used to show slight differences between some structures (i.e., despite and although).”</i>	27
Active participation of students	<i>“Corpus-based activity was like a puzzle for my students. Everyone had an active part in the activity.”</i>	5
Other (s)	<i>“Students in ELT programs can benefit from corpora.”</i>	5
<b>Total</b>		<b>212</b>

N\*= Number of codes

Corpus use in language classrooms was regarded as beneficial due to its *effectiveness in error correction* (69 codes). Participants reported that corpus-based activities might work well in error correction since observing corpus data and identifying accurate uses of the language by themselves could raise students' awareness of their mistakes. The following excerpt exemplifies the instructors' views about the effectiveness of corpus use in error correction:

*T10: "We have used COCA for different purposes, but I believe that one of the most effective functions is error correction. It is possible to use it easily for each level and I believe that I will try to use it to raise my students' awareness."*

As T10 stated above, corpus use for error correction was promising for future practices, and it might also apply to lower levels. Based on their hands-on experiences, the instructors reported that corpus-based activities on error correction engaged students in tasks as they made efforts to find their mistakes and their accurate versions.

The instructors had positive views on corpus use in language classrooms due to its *usefulness in vocabulary instruction* (53 codes). The participants acknowledged that corpora might be useful for teaching particular lexical items where students had difficulties, such as collocations, prefix-suffixes, the knowledge of parts of speech, and confusing words. The following excerpt illustrates the instructors' views about the usefulness of corpus in vocabulary instruction:

*T5: "Especially, to teach confusing words, corpus provides plenty of original sentences in which students can see different uses of those words. By giving tasks in which students analyze the related concordance lines, we can make them aware of the uses of these words and how to differentiate them."*

As T5 noted, corpora might be exploited to introduce word combinations (i.e., collocations), the knowledge of parts of speech, near-synonyms, and confusing words with plentiful samples in various contexts rather than presenting them in isolation. Instructors' views on corpus use were positive due to its *helpfulness for accessing authentic/reference sources* (53 codes); that is, they could utilize corpora as reference tools to check the uses of the language, and they might take advantage of corpora to reach a vast amount of authentic data to design instructional materials. The following excerpt illuminates how corpora can help access to the authentic source:

*T6: "It is an incredibly useful resource for teachers because we are not native speakers and we may not be sure about some uses of the language. So, the corpus can be used to improve our knowledge of the language and to bring reliable materials to the classroom."*

T6 stressed the potential use of corpus tools to observe and/or check accurate uses of the language in actual contexts when they felt uncertain about the use of the English language as non-native speakers of English. It was also noted that corpus might be exploited in preparing classroom materials and creating test items due to its easy accessibility to authentic samples of the language in various contexts. Another benefit of corpus use in language pedagogy was its *usefulness in teaching particular structures* (27 codes). The participants acknowledged that using corpus-based activities might be promising for teaching particular structures (e.g., used

to, gerund/infinitive, and despite/although). Moreover, using corpora in language pedagogy was considered advantageous as it increased *students' active participation* (5 codes) and contributed to some other matters, such as promoting autonomy and appropriateness for ELT students (5 codes).

### **Drawbacks of Using Corpora in Language Teaching**

The final category consisted of five sub-categories regarding the drawbacks of using corpora in language teaching from the instructors' perspectives. The sub-categories are displayed in Table 5.

**Table 5.** Distribution of Codes Related to Drawbacks of Using Corpora

<b>Sub-categories</b>	<b>Explanatory statements</b>	<b>N*</b>
Difficulties in material preparation	<i>"It was difficult and time-consuming to find appropriate concordance lines."</i>	58
Difficulties in students' interaction with corpora	<i>"Students are supposed to have a good level of English to use corpus independently."</i>	48
Impracticality with a low proficiency level	<i>"At lower levels, it is not practical to use corpus due to the challenges."</i>	44
Technical issues	<i>"It was a bit frustrating to go back to the main page and write the same word over and over again."</i>	34
Concerns about grammar instruction	<i>"I am not sure yet if using corpus for teaching grammar is efficient or not."</i>	27
<b>Total</b>		<b>211</b>

N\*= Number of codes

One of the most frequent challenges the participants faced was the *difficulties in material preparation* (58 codes). For instance, scanning concordance lines to find appropriate ones for students' proficiency level and the teaching context was highly demanding. Another concern was the potential *difficulties in students' interaction with corpora* (48 codes). Owing to the nature of corpus data, students would have difficulties in analyzing a great number of concordance lines and making accurate inferences on their own. Hence, *impracticality with a low proficiency level* (44 codes) appeared as another sub-category of challenges. The instructors underlined that implementing corpus-based activities at lower levels (i.e., beginner and elementary) was impractical considering the time and effort required for material design. The following excerpts illuminate the potential difficulties in students' direct interaction with corpora.

*T2: "I think it's difficult for students to use corpus on their own. As we talked in the last session, students might see some wrong usage of the language. It's extremely difficult to use with all students."*

*T4: "I can use COCA to find examples in various contexts at higher levels, but at lower levels, I won't prefer to use COCA, particularly for grammar and the meanings of words because of the challenges in using COCA...The time spent is not equal to the product in the end."*

The instructors also faced some challenges related to *technical issues* (34 codes). Despite having positive attitudes towards COCA in the session, it allowed making a limited

number of corpora queries a day and showed constant warnings to upgrade the user version. Another problem expressed was the difficulties with the layout of materials since it was not so easy to align concordance lines in the Word file. The final sub-category about the drawbacks of using corpora in language teaching indicated some *concerns about grammar instruction* (27 codes). Despite its effectiveness in vocabulary instruction and error correction, the participants were skeptical about the usefulness of corpus-based activities in grammar teaching. The following excerpt shows instructors' concerns about grammar instruction:

*T9: "Although you provided us with different types of sample tasks and tried to inspire us in this way, I still cannot convince myself that it's a good idea to use corpus in grammar lessons. I see that there are some ways to integrate corpus into a grammar lesson, but is it worth the time and effort spent for? I am not sure about it to be honest."*

As T9 stated, the participants had some doubts about the idea of integrating corpus into grammar instruction considering the great efforts made for material preparation. All in all, integrating corpus into language learning through a systematic training program delivered fruitful results in general albeit with certain difficulties.

### Discussion

The results showed that in-service training on corpus use contributed to the instructors' professional development. That is, the participants acknowledged gaining awareness of corpora, becoming familiar with the pedagogical applications of corpora, and acquiring hands-on experiences in corpus-based material design. The instructors reported being able to take advantage of corpora in language classrooms in particular ways together with salient challenges and concerns about corpus use in language pedagogy.

As suggested in the relevant literature, this study implies the potential benefit of creating opportunities to train in-service teachers on the pedagogical uses of corpus tools in the classroom (e.g., Abdel Latif, 2021; Breyer, 2009; Ebrahimi & Faghih, 2017; Farr, 2017; Lin, 2019; Zareva, 2017). In this respect, the results concur with previous studies which pointed out the importance of introducing pedagogical uses of corpora as part of in-service training to influence teachers' classroom practices (Chen et al., 2019; Ebrahimi & Faghih, 2017; Leńko-Szymańska, 2017; Lin, 2019). Constant guidance and support offered in training sessions and hands-on experience provided during the training were prominent aspects of this study. These key features of the training in the present study support the results of previous research regarding the potential benefit of acquiring technical skills to use corpus tools comfortably and effectively (Ebrahimi & Faghih, 2017; Frankenberg-Garcia, 2012; Leńko-Szymańska, 2014, 2017). In a similar vein to Özbay and Kayaoğlu's (2015) study in a Turkish EFL context, this study further highlighted the role and importance of guided training on the use of corpora in actual teaching contexts. One of the striking facets of the current study is that this training program created opportunities for the instructors to actively participate in the process of corpus-based material preparation and implement one of these activities in their teaching context. Concerning teachers' perceived gains from the training, providing the participants with sample materials/tasks, which were prepared specifically based on students' needs in the context of this study, might help the instructors illustrate how to apply corpus-based materials to their

classrooms. Similarly, Poole (2022) pointed out the importance of using corpus-based activities that were designed particularly for the objectives of relevant teaching contexts as teachers and students were required to follow course syllabuses and assignments. Hence, this study builds on the findings of the previous research and adds further light on the relevant literature by taking the needs and requirements of the teachers in their contexts and assisting them in implementing corpus-based materials and activities that reflect the realities of their classrooms.

Apart from teachers' perceived gains, the instructors' perceptions of corpus use in language pedagogy were mainly positive since using corpora as part of language teaching pedagogy was perceived as useful in multiple ways. Corpus-based activities were considered effective in correcting students' errors in writing papers at all language proficiency levels, and the participants noted that corpus use might raise students' awareness of their mistakes since they were actively involved in the process of analyzing data presented through corpora. As suggested in various studies (e.g., Crosthwaite, 2017; Mueller & Jacobsen, 2016; Quinn, 2015), the idea of integrating corpora into language classrooms for error correction was perceived as useful by the teachers. As for vocabulary instruction, this study revealed that teachers might tend to use corpora to introduce particular lexical items such as word combinations (i.e., collocations), confusing words, near-synonyms, and prefixes/suffixes as the instructors in this study perceived corpus use to be more useful and practical for such lexical items. This perception might be closely related to the fact that students are exposed to a great deal of data in which such lexical items occur and they observe patterns related to the language through corpus-based activities rather than learning the English equivalents of words in their mother tongue. Corpus use might raise students' awareness of identifying parts of speech as well. This insight the participants gained in the present study is in line with the findings of the meta-analysis by Lee et al. (2019) as it also suggests that using corpus-based activities can facilitate acquiring in-depth knowledge about lexical items. As reported by the instructors in this study, corpora might be exploited as reference tools when teachers have queries about the language. Since teachers have access to a vast number of examples of authentic language, they can benefit from corpora while preparing instructional materials. This finding highlights the potential benefits of corpora for EFL contexts; therefore, it might help to solve common problems such as preparing materials with authentic sources, and creating test items (Römer, 2009). As cited by the participants in this study, it can also be asserted that corpus-based activities have the potential to foster students' participation in lessons. When compared to conventional ways of teaching and learning, which are mainly based on teachers' explicit instruction and feedback, corpus-based activities involve students in a process that requires analyzing language data and identifying information about language use. Therefore, such activities encourage students to take part in the learning process actively (Lin & Lee, 2015). All in all, along with the potential benefits of corpora for language learners, this study revealed that implementing corpus-based activities in L2 classrooms has the potential to promote the self-development of teachers to become aware of the problems and solutions related to language learning and use at different levels.

Despite all the benefits, problems/challenges related to difficulties in material design, students' direct interaction with corpora, impracticality with lower levels, technical issues, and concerns about grammar instruction were noted as well. As reported by the participants,

rigorous time and effort were needed to design corpus-based activities. This finding echoed similar concerns stated previously in other studies (e.g., Ebrahimi & Faghih, 2017; Lin, 2019). Another drawback of corpus use in language pedagogy was the potential difficulties that might occur in the case of students' interaction with corpora directly. Independent endeavors of corpus use require making accurate inferences about language patterns and having more autonomous and proficient language learners with a certain level of grammar and vocabulary knowledge. Hence, the classroom teacher's guidance and well-preparedness are necessary to present corpus use to language learners who are not familiar with such a pedagogical approach. Ebrahimi and Faghih (2017) also pointed out the skepticism about the applicability of students' direct interaction with corpus tools from the perspective of pre-service teachers.

In this respect, level appropriateness is another issue to consider when using corpus-based activities with lower-level students. Guiding students at lower levels might be hard since they have difficulty interpreting concordance lines despite the presence of the teacher. Such concerns related to the appropriateness of corpus-based activities for lower levels concur with the findings of studies in the relevant literature (e.g., Breyer, 2009; Heather & Helt, 2012). In a more recent study by Poole (2022), two instructors also stated the difficulty of using corpus due to students' lower language proficiency. As noted by the instructors in this study, concerning practicality and suitability, basic vocabulary and grammar knowledge at lower levels may be introduced through textbook activities, teacher-made exercises, and online dictionaries before students are ready to deal with corpus-induced materials.

In line with previous research (e.g., Breyer, 2009; Ebrahimi & Faghih, 2017; Farr, 2008; Zareva, 2017), problems related to the use of technology are stressed in the present study. The participants in this study reported difficulties in using corpus tools and technology in general while designing the layout of corpus-based activities. Considering efficiency and practicality, the idea of integrating corpus into grammar instruction was not welcomed by some of the participants. These instructors asserted that using corpus-based activities might not result in substantial differences related to students' gains in learning grammatical structures, particularly at lower levels when compared to textbook materials. In one experimental study, Lin and Lee (2019) found that there were no significant differences regarding students' gains in learning target structures when compared to the traditional way of teaching grammar in the Taiwanese EFL context. However, the perceptions towards this new treatment were positive over time. With respect to grammar instruction, target structures selected for corpus-based activities can influence teachers' perceptions. The instructors who are teaching at lower levels may face challenges while selecting a grammar structure to adapt to corpus-based activities, especially at lower levels. Lin (2019, p.78) implied that the grammar structures that seem to be more "formulaic" might be better suited for the use of corpus. One reason for some teachers' reported difficulties with corpus tools specifically in grammar instruction might stem from the gaps in their technological knowledge. Such a gap brings forward questions about the problems that may arise from the detachment of the established and familiar teaching methodologies with the use of novel technologies such as the corpora. However, this requires further research and discussion and is not within the scope of the current study. As a result, this study underscores that along with the aforementioned difficulties, practitioners may need more hands-on

classroom practice to gain deeper insights into the feasibility of corpus-based activities in language teaching.

### **Conclusion**

The findings of the study reported here highlight that there is certainly room for integrating corpora into language teaching. That is, teachers might use corpora as one of the teaching aids for classroom activities in teaching particular language items and/or skills. Although numerous gains were perceived in this study, teachers' preferences on the use of corpus-based activities might need to be considered regarding student profiles, learning outcomes, target items and/or structures, and the nature of teaching contexts. It is clearly seen that in-service training plays a vital role in raising teachers' awareness of corpora and equips them with practical skills to use corpus in language pedagogy. Therefore, efforts of meticulously designed training programs can assist in bringing corpus-based activities to actual teaching and learning contexts.

One implication of this study is that corpus use needs to be incorporated into pre-service teacher education programs to empower teacher candidates with the necessary knowledge and skills for their future practices. Another suggestion is that context-specific workshops or training programs might be designed to raise in-service teachers' awareness of the accessibility of corpus tools and how to make use of corpora in their own school contexts. It is highly important to provide constant support and sufficient guidance throughout training to ease teachers' concerns about new applications. Accordingly, teachers can be encouraged to work with their colleagues collaboratively to prepare and implement corpus-based activities. In this regard, teacher educators play a pivotal role in guiding teachers both at pre-service and in-service levels to utilize corpus tools and build a community among teachers that would establish stronger bonds in dealing with the challenges. To this end, teacher educators must become more familiar with corpus tools available for teachers to use in language teaching and learning activities. Considering teacher educators' contribution to teachers' practices in classrooms, teacher educators need to be trained on corpus-based language pedagogy along with all sides including benefits and challenges. This will not only promote the quality of teacher training programs but also will contribute to the quality of the teacher education which will enhance the implementation of corpus-based pedagogy with effective practices. In this way, teacher educators may provide comprehensive training programs to pre-service and in-service teachers that apply to actual teaching contexts. As for textbook writers, corpus-informed textbooks might be popularized to make teachers familiar with the information corpora would offer. Considering teachers' workload, ready-made materials published as activity books might promote the use of corpus-based activities. Apart from these, it is necessary to develop practical tools to facilitate the integration of corpus-based activities into language classrooms.

This study is limited to the views of the participants on their own hands-on practices related to using corpora in English language teaching. Hence, the results cannot be generalized to all EFL instructors. The training lasted for seven weeks, and teachers' corpus-based practices were limited for this period. Long-term studies involving more participants and more practice opportunities could have different results. Therefore, further studies that provide teachers with

training opportunities may be carried out to gain in-depth insights into EFL teachers' perceptions of pedagogical applications of corpora. There is also a need for further research that investigates the long-term influences of in-service training programs on teachers' classroom practices to explore the potential outcomes of using corpora in instructional routines.

### References

- Abdel Latif, M.M.M. (2021). Corpus literacy instruction in language teacher education: Investigating Arab EFL student teachers' immediate beliefs and long-term practices. *ReCALL*, 33(1), 34–48. <https://doi.org/10.1017/S0958344020000129>
- Allan, G. (2003). A critique of using grounded theory as a research method. *Electronic Journal of Business Research Methods*, 2(1), 1-10.
- Aşık, A. (2015). The use of corpora in EFL classrooms: What do Turkish EFL teachers know and offer to promote pragmatic competence? *The Asian EFL Journal*, 17(4), 194-221.
- Boulton, A. (2009). Data-driven learning: reasonable fears and rational reassurance. *Indian Journal of Applied Linguistics*, 35(1), 81-106.
- 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>
- 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>
- Braun, S. (2005). From pedagogically relevant corpora to authentic language learning contents. *ReCALL*, 17(1), 47-64. <https://doi.org/10.1017/S0958344005000510>
- 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>
- Chen, L. (2017). Corpus-aided business English collocation pedagogy: An empirical study in Chinese EFL learners. *English Language Teaching*, 10(9), 181-197. <http://doi.org/10.5539/elt.v10n9p181>
- Chen, M., & Flowerdew, J., & Anthony, L. (2019). Introducing in-service English language teachers to data-driven learning for academic writing. *System*, 87, 1-12. <https://doi.org/10.1016/j.system.2019.102148>
- Corbin, J. & Strauss, A. (2015). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (4th ed.). Sage Publications.
- Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Pearson.
- Crosthwaite, P. (2017). Retesting the limits of data-driven learning: feedback and error correction. *Computer Assisted Language Learning*, 30(6), 1-26. <https://doi.org/10.1080/09588221.2017.1312462>
- Daskalovska, N. (2015). Corpus-based versus traditional learning of collocations. *Computer Assisted Language Learning*, 28(2), 190-210. <https://doi.org/10.1080/09588221.2013.803982>
- 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>
- Farr, F. (2008). Evaluating the use of corpus-based instruction in a language teacher education context: Perspectives from the users. *Language Awareness*, 17(1), 25-43. <https://doi.org/10.2167/la414.0>
- Farr, F. and O'Keeffe, A. (2019). Using corpora to analyse language. In S. Walsh and S. Mann (Eds) *Routledge handbook of English language teacher education* (pp. 268-282). Routledge.

- Flowerdew, L. (2015). Data-driven learning and language learning theories. In A. Leńko-Szymańska, & A. Boulton (Eds.). *Multiple affordances of language corpora for data-driven learning* (pp. 15–36). John Benjamins.
- Frankenberg-Garcia, A. (2012). Raising teachers' awareness of corpora. *Language Teaching*, 45(4), 475- 489. <https://doi.org/10.1017/S0261444810000480>
- Gilquin, G & Granger, S. (2010). How can data-driven learning be used in language teaching? In A. O'Keeffe and M. McCarthy (eds), *The Routledge handbook of corpus linguistics* (pp. 359- 370). Routledge.
- Godwin-Jones, R. (2017). Data-informed language learning. *Language Learning & Technology*, 21(3), 9–27. <https://dx.doi.org/10125/44629>
- Heather, J. & Helt, M. (2012). Evaluating corpus literacy training for pre-service language teachers: Six case studies. *Journal of Technology and Teacher Education*, 20(4), 415-440.
- Huang, L. S. (2017). Has corpus-based instruction reached a tipping point?: practical applications and pointers for English-as-an-additional-language teachers. *TESOL Journal*, 8(2), 295-313. <https://doi.org/10.1002/tesj.271>
- Huberman, M., & Miles, M. B. (2002). *The Qualitative Researcher's Companion*. Sage.
- Johns, T. (1991), Should you be persuaded?: Two examples of data-driven learning materials. *English Language Research Journal*, 4,1-16.
- Karras, J.N. (2016). The effects of data-driven learning upon vocabulary acquisition for secondary international school students in Vietnam. *ReCALL*, 28(2), 166–186. <https://doi.org/10.1017/S0958344015000154>
- Kolb, M. S. (2012). Grounded theory and the constant comparative method: Valid research strategies for educators. *Journal of Emerging Trends in Educational Research and Policy Studies*, 3(1), 83-86.
- Lee, S. (2011). Challenges of using corpora in language teaching and learning. *Linguistic Research*, 28(1), 159–178.
- Lee, H., Warschauer, M., & Lee, J. H. (2019). The effects of corpus use on second language vocabulary learning: A multilevel meta-analysis. *Applied Linguistics*, 40(5), 721–753. <https://doi.org/10.1093/applin/amy012>
- Leńko-Szymańska, A. (2014). Is this enough? A qualitative evaluation of the effectiveness of a teacher-training course on the use of corpora in language education. *ReCALL*, 26(2), 260–278. <https://doi.org/10.1017/S095834401400010X>
- Leńko-Szymańska, A., & Boulton, A. (2015). Introduction: data-driven learning in language pedagogy. In A. Leńko-Szymańska, & A. Boulton (Eds.). *Multiple affordances of language corpora for data-driven learning* (pp. 1–14). John Benjamins.
- Leńko-Szymańska, A. (2017). Training teachers in data-driven learning: Tackling the challenge. *Language Learning & Technology*, 21(3), 217–241.
- Li, S. (2017). Using corpora to develop learners' collocational competence. *Language Learning & Technology*, 21(3), 153–171.
- Lin, M. H., & Lee, J. (2015). Data-driven learning: changing the teaching of grammar in EFL classes. *ELT Journal*, 69(3), 264-274. <https://doi.org/10.1093/elt/ccv010>
- Lin, M. H. (2019). Becoming a DDL teacher in English grammar classes: A pilot study. *The Journal of Language Teaching and Learning*, 9(1), 70-82.
- Lin, M. H., & Lee, J. (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>
- 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>

- Luo, Q. (2016). The effects of data-driven learning activities on EFL learners' writing development. *SpringerPlus* 5,(1255), 1-13. <https://doi.org/10.1186/s40064-016-2935-5>
- Mizumoto, A., & Chujo, K. (2015). A meta-analysis of data-driven learning approach in the Japanese EFL classroom. *English Corpus Studies*, 22, 1–18.
- Mueller, C., & Jacobsen, N. (2016). A comparison of the effectiveness of EFL students' use of dictionaries and an online corpus for the enhancement of revision skills. *ReCALL*, 28(1), 3-21. <https://doi.org/10.1017/S0958344015000142>
- Mukherjee, J. (2004). Bridging the gap between applied corpus linguistics and the reality of English language teaching in Germany. In U. Connor and T. Upton (eds) *Applied corpus linguistics: A multi-dimensional perspective* (pp.239-250). Rodopi.
- Nunan, D. (1992). *Research methods in language learning*. Cambridge University Press.
- Özbay, A. Ş. & Kayaoğlu, M. C. (2015). EFL teachers' reflections towards the use of computerized corpora as a teaching tool in their classrooms. *Çukurova Üniversitesi Eğitim Fakültesi Dergisi*, 44(1), 85-104. <https://doi.org/10.14812/uefd.54367>
- Quinn, C. (2015). Training L2 writers to reference corpora as a self-correction tool. *ELT Journal*, 69(2), 165–177. <https://doi.org/10.1093/elt/ccu062>
- Poole, R. (2022): “Corpus can be tricky”: revisiting teacher attitudes towards corpus-aided language learning and teaching, *Computer Assisted Language Learning*, 35(7), 1620-1641. <https://doi.org/10.1080/09588221.2020.1825095>
- Römer, U. (2009). Corpus research and practice. In K. Aijmer (ed.) *Corpora and language teaching* (pp.83-98). John Benjamins.
- Römer, U. (2011). Corpus research applications in second language teaching. *Annual Review of Applied Linguistics*, 37, 205–225. <https://doi.org/10.1017/S0267190511000055>
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11, 129–58.
- Smart, J. (2014). The role of guided induction in paper-based data-driven learning. *ReCALL*, 26, 184–201. <https://doi.org/10.1017/S0958344014000081>
- Timmis, I. (2015). *Corpus linguistics for ELT*. Taylor & Francis.
- 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>
- Zareva, A. (2017). Incorporating corpus literacy skills into TESOL teacher training. *ELT Journal*, 71(1), 69–79. <https://doi.org/10.1093/elt/ccw045>

## Appendices

### Appendix A: Interview questions (prior to the training)

1. Are you familiar with the term “corpus”? If yes, how do you define it?
2. Do you know any corpora? If yes, which ones?
3. Have you ever used corpora? If yes, why? And how?
4. Do you think corpora can be used for English language teaching and learning purposes? If yes, how?

### Appendix B: Sample materials/tasks used in the training

#### A sample task used in Session 2

**Exercise A.** Read the questions below and search them in COCA to find the answers.

1. a. Find the frequency of the lexical bundles below and put them in order (1-3).

(The lexical bundles below are commonly used in students' writing papers in our context.)

- In my opinion \_\_\_\_\_
- I believe that \_\_\_\_\_
- As far as I'm concerned \_\_\_\_\_

1. b. Search the frequency of each lexical bundle below by genre and put them in order (1-5).

<i>In my opinion</i>	<i>As far as I' m concerned</i>	<i>I believe that</i>
<i>Academic</i> _____	<i>Academic</i> _____	<i>Academic</i> _____
<i>Fiction</i> _____	<i>Fiction</i> _____	<i>Fiction</i> _____
<i>Magazine</i> _____	<i>Magazine</i> _____	<i>Magazine</i> _____
<i>Newspaper</i> _____	<i>Newspaper</i> _____	<i>Newspaper</i> _____
<i>Spoken</i> _____	<i>Spoken</i> _____	<i>Spoken</i> _____

2. Follow the instruction below and find the top three nouns collocate with the word “record”.

- Insert “POS” (noun.all) and search for the results. Write them below.

- \_\_\_\_\_ record
- \_\_\_\_\_ record
- \_\_\_\_\_ record

3. Follow the steps below and compare the words “say” and “tell” to find what nouns collocate with them.

- Choose “COMPARE”.
- Write the word “say” as word 1 and write the word “tell” as word 2.
- Insert “POS” (noun.all) in the blank “collocates”.
- Limit your search to the first word on the right.
- Examine the nouns in the top 20 list and write five of them below.

- Say \_\_\_\_\_
- Tell \_\_\_\_\_

### Sample tasks used in Session 3

Exercise A. Read the questions below and search them in COCA to find the answers.

#### 1) question/ wonder

##### ➤ Search for Question\*

- Find two different concordance lines in which the word is used in a different part of speech (e.g., as a noun and as a verb).
- Identify the part of speech in each concordance line that you have found.
  - In the 1st concordance line, the word \_\_\_\_\_ is a/an \_\_\_\_\_.
  - In the 2nd concordance line the word \_\_\_\_\_ is a/an \_\_\_\_\_.

##### ➤ Search for Wonder\*

- Find two different concordance lines in which the word is used in a different part of speech (e.g., as a noun and as a verb).
- Write down the sentences in which the words appear.

\_\_\_\_\_

\_\_\_\_\_

- Generally, the word (as a verb) *wonder* takes the preposition: \_\_\_\_\_

Exercise A. Analyze the concordance lines below and answer the questions below.

#### ➤ Circle the correct word. *Job or work?*

1. ...there are also more fundamental issues related to incomes and job/work opportunities that split California into a two-speed economy.
2. ...said he was troubled by the damage. "I would not job/work with them after that," he said. "I would not refer..."
3. ...health insurance and rent. I saw two options. Get a job/work return my advance and abandon two years' worth of work...
4. ...told an unnamed colleague she feared she would lose her job/work if she reported the incident. Another unnamed person reported Moulton...
5. ...at his home to discuss why he was late to job/work, the station reported. The victim told the sheriff's office...
6. ...grown-up, he said. He's been doing a lot of job/work around the house, finishing the kitchen project his dad started.

- Write the part of speech of the target words above.

\_\_\_\_\_

\_\_\_\_\_

- Look at the lines above. Which word can be used in plural? \_\_\_\_\_

**Exercise A. Analyze the concordance lines below and answer the following questions. They will help you find patterns either in the left or right context (KWIC).**

...less likely . We 'd heard a lot of rumor and gossip about Juliet, but we 'd heard nothing about Pauline , and...  
 ...and their topics of conversation were limited to TV dramas and gossip about their children 's teachers , so she ...  
 ...'s brother - married Hamlet 's mother . There is much gossip about these events , and Hamlet is very upset .  
 ...did gossip about us , and I knew that I 'd gossip about us if I were someone else . For a brief moment...  
 ...younger, but also now, dealing with a lot of gossip and rumors about your life, and how have you coped with...  
 ...got down to work . Raymie is a geyser of gossip and hard news, a Wife of Bath ; she knows people...  
 Her extramarital antics were a source of gossip and shame that affected the entire George family...  
 What were they thinking? It is what make gossip and what makes afternoon and evening TV work, "he said...  
 ...novels . She was never sneaky , for she openly gathered gossip and you could see her editing it in advance with a little  
 ...if you have enough money. TAIBBI: (Voiceover)The Hamptons gossip columnist, who calls himself " Sudsy, " says he did...  
 ...him for a handshake or an autograph. For a local gossip columnist he puts down his Diet Coke and poses for a...  
 ...There they are, smiling and waving, posing with a gossip columnist from - I don't know -Paraguay. Occasionally ,...  
 ...places to juggle, a few bruised were inevitable --setting gossip columns abuzz with rumors of supposed slights and rebuffs.  
 ...a celebrity executive - one whose name frequently shows up in gossip columns. Gossip also is paying off handsomely for his ...  
 ...to 36-year-old supermodel Christina Estrada. It is the stuff of gossip columns. #Lately, he 's been a busy sprite ,...  
 ...his profession and the wife could be unafraid of whispered gossip in the market. # By the end of 157, an...  
 ...I 'll save you a program and tell you all the gossip in the morning. " " Oh, I do wish I...  
 ...to journalism. So closely was Winchell identified with gossip in the late twenties and thirties that many readers thought it...  
 ...to admire their mistresses' long legs in short tunas. Gossip in the dressing room usually centered on ballerinas who had...  
 ...faces to the warming sun." To an Italian, gossip is second only to religion in order of importance. The better...  
 ...while trading voluminous amounts of news and gossip was one of the wonders of ' modern physiology. Yet it...  
 ...hat she would bring a sizable dowry, which meant the gossip was as wrong as always. " Francis ! " Arising...

1. Look at the concordance lines above. Is *gossip* a verb or noun? Which one is more frequently used in this body of data?  
\_\_\_\_\_
2. Look at the right context of the word. Which preposition is commonly used with "*gossip*"?  
\_\_\_\_\_
3. Look at the right context of the word. When is it used with the preposition "*in*"?  
\_\_\_\_\_
4. Which job collocates with the word "*gossip*"?  
\_\_\_\_\_
5. Look at the data. Which nouns frequently collocate with the word "*gossip*"?  
\_\_\_\_\_
6. In what kind of texts/sections do you think you are more likely to find the word "*gossip*"? Put them in order from the most frequent to the least one.

Academic	
Fiction	
Magazine	
Newspaper	
Spoken	

Session 4

A sample corpus-based activity on vocabulary by one of the participants (T5).

Task 1

1. -ous

... But at the same time we have to be cautious about what companies we invest in. # Analysts say more	cautious	about	what	companies	we	invest	in
'd stored. But it was n't himself that he was curious about     was her thoughts, her feelings. # And	curious	about			was	her	thoughts, her feelings.
4 sons and requested them to give a part of their pitous acts for redeeming Yayer. I also offered half of my accumulated	pitous	acts	for	redeeming	Yayer	I	also offered half of my accumulated
Page 1 Times story by Judith Miller and Michael Gordon cites anonymous administration official saying Saddam has repeatedly tried to	anonymous	administration	official	saying	Saddam	has	repeatedly tried to
this testing, 400,000 or so tests which cost us an enormous amount   four hundred to four hundred and fifty actually HIV	enormous	amount		four	hundred	to	four hundred and fifty actually HIV
7 RF: The show was wonderful in Washington; very ambitious and beautifully installed   They worked a long time on it,	ambitious	and	beautifully	installed		They	worked a long time on it,

2. -ity

ho!" There was some laughter and a str of curiosity about our lime craft as it drew closer. # I wondered	curiosity	about	our	lime	craft	as	it drew closer.
workers had become exhausted, cynical, and divided; Most security analysts have been slow to understand this reality and have	security	analysts	have	been	slow	to	understand this reality and have
of attitude and sensibility, allowing for a focus on internal identity and a shifting, complex worldview. # John's explication of his	identity	and	a	shifting,	complex	worldview.	# John's explication of his
criticism from its normative bounds and as an antidote to the aridity and affective repression represented by High Theory (Tomplins	aridity	and	affective	repression	represented	by	High Theory (Tomplins
, Thompson's account seeks to preserve the validity of subjectivity and agency as features or products of 'I'-making.	subjectivity	and	agency	as	features	or	products of 'I'-making.

3. -ion

has enabled the CPPD to portray itself as the force for democratization and the Right as the obstacle in its path -- thereby maintaining	democratization	and	the	Right	as	the	obstacle in its path -- thereby maintaining
narrative. The assessment worker needs the information for the investigation and to decide on placement. " Once a report has been	investigation	and	to	decide	on	placement.	" Once a report has been
a hotel; complete with spaceplane service from dirtside to the station and travel via private LTV and landing module to his own field	station	and	travel	via	private	LTV	and landing module to his own field
his father always told him the mining game was a risly proposition and urged him to go into teaching. It is a lesson	proposition	and	urged	him	to	go	into teaching. It is a lesson
belt across the width of the continent and in a southeastern region around northern Tanzania   Using a computer simulation to	region	around	northern	Tanzania		Using	a computer simulation to

4. -ful

wrote in his memoirs that his young assistant * did a masterful job   disclosing the whole anatomy of Wall Street chicanery and	masterful	job		disclosing	the	whole	anatomy of Wall Street chicanery and
# Henningsen turned towards the exit not noticing the look the beautiful lady   black sent him from the tomb opposite that of his	beautiful	lady		black	sent	him	from the tomb opposite that of his
age. # What Sunny Jim's creation needed was another forceful leader who could turn the 1989 earthquake to City Hall's	forceful	leader	who	could	turn	the	1989 earthquake to City Hall's
Gulf Coast from Los Angeles two years ago seeking a more peaceful life and how finds his beachside cottage 200 yards from an	peaceful	life	and	how	finds	his	beachside cottage 200 yards from an
by being provided artificial opportunities to disclose their stressful life experiences   as found previously in samples of male	stressful	life	experiences		as	found	previously in samples of male

Analyze the concordance lines above and find out which suffixes form words as NOUNS or ADJECTIVES.

Suffix	Part of Speech (N or Adj)	Suffix	Part of Speech (N or Adj)
-ous		-ion	
-ity		-ful	

## A sample task used in Session 5

**Exercise A. Analyze the concordance lines below and circle the correct option.**

...the eastern shores of the Adriatic Sea, the city has been an important seaport **since** the Middle Ages. 44 In 1992, when Yugoslavia...  
 ...to Koistinen in recent times. Indeed, it has been more than a decade **since** it was last cited in a case. See *Marine Solution Servs., Inc. v...*  
 ...of assisted passage by Australia Finnish Immigration petered out in the early 1970's. **Since** then it has mostly been professionals of...  
 ...has instituted major revisions to the process (Steinzor et al, 2009). **Since** 2004, revisions to the IRIS process have given peer reviewers...  
 ...George has experienced an enduring economic decline **since** World War II. As one of Detroit's industrial, inner-ring suburbs, Ecorse...  
 ...she said. # Layney has been in Scouting **since** the first grade, and is the daughter of Rick and Nancy Duncan of Little...  
 ...is doing well at school, mostly. She has known how to read ever **since** she was three. Frances loves to read but struggles to speak, or to...  
 ...to battle the kind of special interest legislation that the Court has tolerated **for** more than forty years. In short, the Supreme Court...  
 ...times. I think there's great respect. We've been friends **for** a long time. # "When he was here at Tulsa, we played them over...  
 ...caught his arm. " My god, I haven't seen you **for** ages. " # Without pausing in his cleaning, keenly aware of the security camera...  
 ...but you have not forgotten what you have done **for** many years and for many hours every week using an instrument...  
 ...They're from Holland, Michigan. They've been married **for** three years. She's a teacher on school break. AL ROKER: And couple...  
 ...husband was taking a professional risk. Now that you've been there **for** six months and people are well aware of your presence there...

1. Which tense is used with *since/for* in the given examples above?

2. We use *since* ...

- to talk about a period of time (how long)
- to talk about a point in time (when something started)

3. We use *for* ...

- to talk about a period of time (how long)
- to talk about a point in time (when something started)

## Session 6

A sample corpus-based activity on grammar by one of the participants (T9).

**EXERCISE A:** Analyze the concordance lines below and answer the following questions.

18465	How much	adventure can a ski boot handle ? We explored the possibilities
13117	How much	protection are you really getting? Not much . (Fully 42%
of the total party sales . RAZ : So ↓	how much	money did you raise? Ms-NEEDHAM : We raised \$450 that weekend
to get the money - economy moving quickly ? And	how much	money did you spend on the long-term aspects of rebuilding an
so much for being with us this morning . So	how much	truth do you think there is to some of the comments in

with a glittering eye . And I say , ↑	How many	hours a day do you work ? ' And he says ,
his eyes . This fellow was smooth . # "	How many	tickets does Blaine need to sell ? " he asked . The
lot more intense , I believe . ZAHN : And	how many	questions are you getting peppered with from -- from many of your
it ? He also asks her ... UNIDENTIFIED MALE :	How many	found can you last? UNIDENTIFIED FEMALE : I dont know.
a courier ? Woman 2 : Yes . GOLDBERG :	How many	trips did you make for the Nigerians ? Woman 2 : Just

- Which one do we use with uncountable nouns; **how many** or **how much**?
- Which one do we use with countable nouns; **how many** or **how much**?
- With how many, do we use the **plural form** or **singular form** of countable nouns?
- When we ask questions by using "how many/ how much", the word order is

### A sample task used in Session 7

**Exercise A.** Look at the concordance lines retrieved from COCA. Correct the mistakes in the following sentences.

1.

...looking for someone to oversee or keep an eye on her because we're all making the film.' That guy Bobby...  
 ...he can't drink coffee for a quick jolt during the game because it might affect his stomach, so he has turned to...  
 ...what makes this category fun and useful. Because it weighs less than half a pound, you don't have to...  
 ...for gathering accurate polling information, in large part because anyone can read your tweet and reply to it.  
 ...a credit, Philadelphia can minimize its loss of revenue because the credit will only be applicable when a...

It is necessary for some websites. Because some people can share illegal content on social media.

2.

...when politicians -- presidential candidates go abroad, they have meetings with local leaders...  
 ...to get out fast # If you simply can't wait to go abroad, look into jobs that will take you out of the country...  
 ...mainly to learn English, since they planned to go abroad for professional school or graduate work.  
 ... students stayed at their countries and did not go abroad, they would not be able to have...  
 ...our growing economy, it is now easy for us to go abroad for studies or work, thereby making it possible for us...

You can improve your English when you go to abroad.

3.

...social historian would choose to scale twice. This connects to a dilemma for any movie lover -- the choice about  
 ...researcher for TACC and lead author on the paper. "This gives radiologists and other clinical staff the means to...  
 ...real dog's life in danger while filming the movie. This comes from TMZ (via Consequence Of Sound), ...  
 ...to rotate the rocket between 120 and 180 degrees. This aims the rocket toward the drone ship. The booster is...  
 ... for about 15 minutes on low to medium heat. This relaxes the fabric and releases odors. # Prevent wrinkles....

This cause air pollution.

4.

...through photographs and text children are shown how to do experiments that explore different kinds of energy.  
 ...to remove negative ideas from the brain. We've done experiments on monkeys, but never on humans.  
 ...using our oceans in the Gulf of Mexico for that. We're doing experiments there now, research. There's so much but all  
 ...take all the safety precautions?... # " If you're going to do experiments, fine. But don't allow people to live in and

Scientists shouldn't make experiments on animals.

### Session 8

A sample corpus-based activity on error correction for intermediate level (T2).

Look at the concordance lines retrieved from COCA. Correct the mistakes in the following sentences.

1. What is the part of speech of “affect”? \_\_\_\_\_  
What is the part of speech of “effect”? \_\_\_\_\_

It is clear that it effects relations badly.

.....how the confirmation of Betsy DeVos as education secretary might **affect students** and teachers across the country.....

.....the generalization passage did not **affect students'** reading of high-frequency words. There.....

..... with regard to understanding when and how learning disabilities **affect students'** academic performance and overall social behavior. Moreover,.....

These time variables are proxies for the net **effect of** all the economic factors that create.....

.....it does have a clear negative **effect on** the management fee load. Hence, the .....

.....of young adults, consistent with the negative **effect on** childhood mathematical achievement.....

*A sample corpus-based activity on error correction for elementary level (T10).*

**Part A. Look at the concordance lines retrieve from COCa. Correct the mistakes in the following sentences.**

...those men have to go home and **do their homework** and talk about what attacks could happen...  
... on the bus and I use the time to **do my homework** and read. Mostly boks about ...  
... this would be the first time I didn't **do my homework!** # That night at dinner, I bring up...

I make my homework every day at 7 pm.

... " I'll take care of Jane Doe. You **go home** and get some sleep. Save some up for after...  
... the night. If the weather improves, you can **go home** in the morning. " # Priscilla...  
... , do you? # " Come on, Dad, let's **go home**, " Leland said. # Sherm shook his head. " Not until...

My lesson finishes at 4 in the evening and I go to home.

### *A sample task used in Session 8*

- Design a corpus-based material for the level you're teaching in one of the areas covered in the sessions. Take notes based on the following elements.
  - Language proficiency level  
\_\_\_\_\_
  - Aim of the activity  
\_\_\_\_\_
  - Target lexical items/ structures / error correction  
\_\_\_\_\_
  - Genre (If any, e.g. spoken, academic, newspapers, etc.)  
\_\_\_\_\_
  - Activity type (e.g. gap filling, multiple choice, open-ended, etc.)  
\_\_\_\_\_
  - Activity format (printout materials vs. computer based version)
  
- Implement the material you've prepared in the classroom (If applicable, you can use one of the materials you prepared in the previous sessions).

## Appendix C: Follow-up interview questions

1. What do you think about workshop sessions?
2. What do you think about COCA as the tool?
3. Do you think that teachers and/or learners may profit from corpus data for language teaching and learning? If yes, for which language areas and how?
4. In the future, would you consider consulting any corpora for language teaching? If yes, how would you use it?