

Yabancı Dil Öğretimi Sınıflarında Derlem Kullanımı Eğitimi: Öğretmen Adaylarının Algıları

Berkcan KÖSE^{a, b}, Safiye İpek KURU GÖNEN^c

Özet

Bu çalışma, Türkiye'de bir devlet üniversitesinde öğrenim gören 108 Türk İngilizce öğretmeni adayının yürütülen iki haftalık derlem eğitimi ardından derlem kullanımının dil öğretimine entegre edilmesine ilişkin algılarını araştırmıştır. Derlem kullanımının dil öğretiminde kullanımına yönelik artan ilgiye rağmen, çok az araştırma İngilizce öğretmeni adaylarının derlem kullanımına ilişkin algılarını ele almıştır. Nitel araştırma yönteminin uygulandığı bu çalışmada, demografik bilgi ve derlem kullanım anketi, yarı yapılandırılmış görüşmeler ve yansıtma raporları gibi çeşitli veri toplama araçları kullanılmıştır. Sonuç olarak, öğretmen adaylarının derlem kullanımı kelime ve dilbilgisi öğretiminde ve öğretim materyalleri hazırlamada faydalı bulunduğu tespit edilmiştir. Derlem kullanırken karşılaştıkları çeşitli zorluklara rağmen katılımcıların gelecekteki öğretmenlik deneyimlerinde derlem araçlarını kullanmak istedikleri ortaya çıkmıştır. Bu bulgular, öğrencilerinin dil öğrenme deneyimlerini zenginleştirmek için sınıflarında derlem kullanmayı tercih eden öğretmen yetiştiricileri ve öğretmenler için yararlı olacaktır.

Anahtar Kelimeler

Dil Öğretiminde Derlem
Kullanımı
Yabancı Dil Öğretimi
İngilizce Öğretmen Adayları
Öğretmen Eğitimi

Makale Hakkında

Geliş Tarihi: 03.06.2023
Yayın Tarihi: 28.12.2023
Doi: 10.18026/cbayarsos.1309293

Training on Using Corpora in EFL Classrooms: Perceptions of Pre-Service Teachers

Abstract

This study investigated the perceptions of 108 Turkish pre-service EFL teachers (PSTs) at a state university in Turkey regarding integrating corpora into language classrooms after the two-week training conducted. Despite the growing interest in using corpora in EFL contexts, very little research has addressed pre-service EFL teachers' perceptions of corpus use. Following a qualitative research design, multiple data collection tools, such as a background survey, pre- and post-training semi-structured interviews, and reflective logs, were utilized. The findings revealed that the PSTs found corpus use beneficial for teaching vocabulary and grammar and preparing authentic teaching materials despite a few drawbacks, such as providing limited daily access and having too much data to analyze. Their remarks revealed that the PSTs would consult corpora in their future teaching experiences. These findings would be useful for teacher educators and teachers who favor using corpora to enrich language learning experiences.

Keywords

Corpus Use in Language
Teaching
Foreign Language Teaching
Pre-service Language Teachers
Teacher Training

About Article

Received: 03.06.2023
Published: 28.12.2023
Doi: 10.18026/cbayarsos. 130929

^a Correspondent Author: berkcan@metu.edu.tr

^b Res. Asst., Middle East Technical University, Faculty of Education, Department of Foreign Language Education, ORCID: 0000-0001-5414-7583

^c Assoc. Prof. Dr., Anadolu University, Faculty of Education, Department of Foreign Language Education, ORCID: 0000-0002-7374-8269

Introduction

The dynamic nature of English as a Foreign Language (EFL) teaching has been deeply affected by the recent innovations in computer sciences, the integration of technology into L2 classrooms, and the discovery of ways to benefit from what is available online with one click (Jamal, Shafqat & Afzal, 2021). Therefore, the use and integration of online tools into language classrooms and computer-assisted language learning (CALL) applications for teaching instruction and material design have been gaining prominence in L2 research (Alsolami & Alharbi, 2020).

As one of CALL's most outstanding applications, there has been a growing interest in the use of corpora and corpus-based teaching pedagogy in L2 contexts recently (Boulton & Cobb, 2017; Crosthwaite, Luciana & Schweinberger, 2021; Forti, 2019). Corpus is simply the electronic collection of authentic language samples of written and spoken texts in various contexts and genres (Biber, Conrad & Reppen, 1998; Hunston, 2002; Johns, 1991). With the help of corpus consultation, teachers and learners are provided with various features and aspects of word knowledge regarding word frequency in formal/informal contexts in different years, as well as the patterns in the language regarding collocations, idioms, and lexicogrammatical usage (Mizumoto & Chujo, 2015). Corpus use has been widespread since it provides natural and authentic language samples, and it also enhances learner autonomy since learners can be the 'researchers' themselves while searching for a specific item in the corpora and exploring the presence of authentic use of language patterns through concordances (Çalışkan & Kuru Gönen, 2018; Yoon & Hirvela, 2004).

Although several previous studies explored the perceptions and attitudes of in-service teachers and EFL learners, very little research has addressed pre-service EFL teachers' perceptions regarding integrating corpora into language classrooms (Crosthwaite et al., 2021; Leńko-Szymańska, 2017). Moreover, ELT programs in the Turkish context neglect the advantages of corpora consultation and provide little or no attention to the corpus-based language pedagogy in pre-service level education. Therefore, this present study is significant in that the pre-service EFL teachers were given training based on using corpora and preparing corpus-based teaching materials. Their perceptions regarding the use of corpora and corpus-based teaching materials were investigated. This study serves as a piloting to develop a comprehensive corpus training model for pre-service teacher education; hence, the results will shed light on how to integrate corpus-based language pedagogy in English language teaching contexts.

Consulting a Corpus in EFL Teaching

A corpus, known as the online archives of authentic language samples, including written and spoken input, provides a technology-based language learning and teaching methodology that has had a tremendous impact on the shift to a more technology-based philosophy in language teaching contexts and applied linguistics (Alsolami & Alharbi, 2020; Biber et al., 1998). Römer (2010) also suggests that corpus use for language teaching is of two types: indirect use by researchers and material writers and direct use by teachers and learners for learning/teaching purposes, also known as *data-driven learning* (DDL), coined by Johns (1991). DDL is considered the most direct and proper application of corpora, as it has been used for over 30 years. With

the help of DDL, learners are given chances to analyze authentic examples of usage patterns of vocabulary and grammatical features (Mizumoto & Chujo, 2015).

The extent to which teachers and learners can benefit from the integration of corpora into language learning contexts depends on their corpus literacy regarding the corpus queries for finding keywords in context and analyzing concordance lines, their familiarity with corpus tools and online corpora, problems related to the access online corpora and the lack of instructions and training related to corpora use, and most significantly, learners' and teachers' experiences, perceptions, beliefs, and willingness towards the use of corpora and DDL tasks (Farr, 2008). According to the current research on the perceptions of teachers regarding corpus use, DDL is believed to raise the awareness of learners towards the recognition of word patterns and collocations, enhance the teaching of grammar and vocabulary items, develop learners' error correction and proofreading skills, and emphasize learner autonomy (Cobb & Boulton, 2015). As well as the advantages of corpus consultation by teachers, a wide range of drawbacks of DDL have been stated in the current research regarding its use, such as the complexity of the interfaces of the corpus tools and online corpora, negative perceptions and attitudes of learners and teachers towards the corpus use, and the difficulty of its integration into language classrooms (Gilquin & Granger, 2010; Liu & Jiang, 2009; Rasikawati, 2019).

Özbay and Kayaoğlu (2015) examined the perceptions of six EFL teachers at a university-level prep school who had "little or no contact" with corpora and conducted corpus training. The findings pointed out that the participants' perceptions of corpus use were positive and favored using technological tools in their teachings. In a similar study, Çalışkan and Kuru Gönen (2018) aimed to investigate the effect of the corpus on vocabulary teaching and the perceptions of three EFL instructors at a state university in Turkey. It was revealed that the participant teachers had no prior contact or experience using corpus to teach vocabulary. As a result of the training they received, they were content with teaching vocabulary via corpus-based materials.

Zareva (2017) also investigated the attitudes and beliefs of 21 MA TESOL teacher trainees toward utilizing a corpus. The participants' perceptions were generally positive since they felt as if they had been discovering a new world. However, they also experienced challenges regarding the technical aspects and the interface of corpora and perceived analyzing corpus data as time-consuming. More recently, Jamal et al. (2021) examined the perceptions of 19 EFL instructors toward corpus-based teaching. The participants benefited from consulting corpora for teaching skills as they found it student-centered and believed it enables learners to decide for themselves. In her MA thesis, Çalışkan (2020) revealed that 89 EFL instructors' familiarity and experience regarding corpus use had been low or non-existent. Following the instruction, they found it useful for error correction and teaching vocabulary. However, they stated that preparing materials with corpora is complex, and low-level learners might have problems in a corpus-assisted language teaching environment.

Recently, Bal-Gezegin, Akbaş, and Başal (2022) aimed to explore how pre-service language teachers viewed the consultation of corpus in providing corrective feedback. In this qualitative research, a cohort of pre-service language teachers was assigned written texts for an error correction exercise, utilizing a corpus to assist them during the exercise. The findings revealed that pre-service language teachers preferred using corpora and were eager to integrate them into their English learning and teaching. The results demonstrate that corpus consultation benefits not only their feedback-giving abilities but also enhances their autonomy, awareness,

and motivation as learners. Poole (2022) also examined the viewpoints of novice TESOL instructors on corpus-based language teaching in a US university's undergraduate writing course for multilingual students. While the instructors acknowledged the value of corpus activities, they expressed concerns about the effectiveness of pre-made activities in promoting independent student inquiry. Despite students' mastering search functions and completing activities, they didn't engage in the expected discovery-based, inductive learning associated with corpus study.

In a similar vein, Lin and Lee (2015) aimed to investigate the perceptions of six novice EFL teachers who stated that DDL tasks were beneficial for grammar teaching and better-engaging learners in their learning process. The findings revealed several drawbacks of DDL, such as the workload for teachers and learners and the technical problems related to the use of online corpora. Chen, Flowerdew, and Anthony (2019) also investigated the impact of a teacher training workshop on DDL and the perceptions of 54 in-service EFL teachers. The findings revealed that they favored DDL tasks, and the challenges they experienced were related to their prior experience and knowledge about corpora.

In the pre-service EFL context, Crosthwaite et al. (2021) investigated the perceptions of 22 pre-service teacher trainees. The participants favored the corpus training as they developed their corpora and stated that corpus use is beneficial for learning grammatical features and vocabulary. Leńko-Szymańska (2017) also implemented a semester-long training on corpora use in language teaching as a course. The study revealed that the participants could not meet the competencies desired for the successful use of corpora and lacked pedagogical skills and autonomy. The studies conducted with PSTs suggested the need for further studies and more systematic training for the successful implementation and consultation of corpus use into EFL teaching and a training framework in pre- and in-service teacher education.

Significance of the Study and the Research Questions

Although the majority of the studies mentioned above aimed at exploring the perceptions and attitudes of in-service teachers and EFL learners, very little research has addressed pre-service EFL teachers' perceptions regarding the integration of corpus and DDL tasks into language classrooms (Crosthwaite et al., 2021; Leńko-Szymańska, 2017). It is also well-known that ELT programs in the Turkish context neglect the advantages of corpus consultation and provide little or no attention to the corpus-based language pedagogy in pre-service level education. Thus, this present study is significant in that the pre-service EFL teachers were given training based on using corpora and preparing corpus-based teaching materials. Their perceptions regarding the use of corpora and corpus-based teaching materials were investigated. The following two research questions were formulated based on the purposes of the study:

- a. Are Turkish pre-service EFL teachers familiar with corpus tools and their use in EFL classrooms? If yes, to what extent?
- b. What are the perceptions of Turkish pre-service EFL teachers regarding the integration of corpora and DDL tasks into language classrooms following the corpus training?

This study will propose several implications for integrating corpus into teacher-training programs and provide insight into the actual classroom applications of corpus-based language pedagogy.

Method

Context and Participants

The current study was carried out at the Foreign Language Education (FLE) Department of the Middle East Technical University (METU) in Ankara, Turkey, in the 2021-2022 Spring Term. A total of 108 sophomore students enrolled in the four-year English Language Teaching program were enrolled in the study. Consecutive two one-hour training sessions were conducted with the participants over two weeks as a part of the course syllabus, with no extra credits or grades for their attendance. Researchers obtained each participant's informed consent and the department's ethical approval. The second-year sophomore students of the FLE Department were considered PSTs in this study as they had been offered methodology courses that had provided them with chances of micro-teaching. The majority of the group had had EFL teaching experience of at least half a year either as an EFL tutor or as an EFL teacher at private schools before the study. A convenience sampling strategy (Creswell, 2012) was followed as all the participants were volunteered, accessible, and willing to participate in the study. All 108 PSTs attended the training sessions, and qualitative data were collected via various data collection instruments. A background survey was administered before the training provided that the participants were between 18-25 years old, and they had varying teaching experience tutoring students from elementary to graduate lessons, mostly within private courses. All PSTs were involved in micro-teaching experiences in various ELT-related methodology courses, teaching to their peers as part of the program's requirements. The participants received no specific course or training regarding corpus use. Hence, in this study, all PSTs received training on corpus use for the first time.

Training on Corpus Use and Corpus-Based Materials

Two one-hour training sessions were conducted over two weeks with the participants. While the first sessions were conducted in the classrooms, the second training sessions were conducted in the department's computer lab so that the participants would have individual computers to access the tasks and the interface of the Corpus of Contemporary American English (COCA), the corpus tool utilized in this study. The researchers designed the content of the training sessions in line with the purposes above and the research questions of the study and presented various presentations, discussions, tasks, and one in-class assignment to provide hands-on experience of COCA to the participants and raise their awareness on corpus consultation for language teaching purposes. The content of the training sessions regarding 3 presentations, 15 tasks, 1 in-class assignment of preparing a teaching material by consulting corpus, and discussion questions was adapted from various books and studies in the corpus-based research literature (Baker, Hardie & McEnery, 2006; Bennett, 2010; Çalışkan, 2020; Leńko-Szymańska, 2015; O'Keeffe, McCarthy & Carter, 2007; Özbay & Kayaođlu, 2015). The researchers consulted four experts in the field of ELT and corpus pedagogy to ensure the reliability and validity of all content of the training sessions, presentations, and tasks.

In the *first session* of the training, the learners were introduced to the basic notions and types of corpora, what a corpus provides, the ways and benefits of consulting corpora and concordance lines through pedagogical corpus applications, and the concept of selection of vocabulary and grammar teaching items by consulting corpora. Figure 1 below shows several sample tasks used in the first session.

What does a corpus provide?

- Frequency of words
- Genre (written or spoken)
- Collocations, phrasal verbs, idioms
- Meaning
- The use of modality and tenses
- Context
- Register (formal vs. informal)
- **Vocabulary size**

Task 3

What do you think are the top three nouns that collocate with the verb "make"?

make

make

make

make

make

Task 4

Figure 8: Extract from *English Idioms in Use* (McCarthy and O'Dell 2002: 109)

50.4 Here are some random examples from a computer database containing lines from real conversations. The figures in diamond brackets, e.g. <s1>, <s2>, mean 'first speaker', 'second speaker', etc. How many of the examples use *eye* as an idiom, and how many use the word *eye* in its literal sense as 'the organ we see with'? Use a dictionary if necessary.

```
1      go into town and get erm an eye test. <s1> Mm. <s2> > In town.
2              you er keep an eye out for tramps, do you then?
3              In your mind's eye how are you going to do that?
4      <s1> So I'll keep a general eye on it. And er <s3> Yeah
5      <s1> There's something in my eye. There's that thing floating
6              difficult to put that to your eye. You also have to have one eye
7      good offer? <s2> Yeah it caught my eye <s1> Yeah it's
8              I'm casting my eye over this form and I think
9              this year. <s4> Just keep an eye out for it. <s4> Yeah.
10     <s2> You'll have to keep an eye on her. <s1> Yeah. <s2> Oh my
11              so you're about eye level with the monitor.
12     saw her out of the corner of my eye. <s3> Her lipstick is all over
```


1. How many of the examples use 'eye' as an idiom?
2. How many of the examples use the word 'eye' in the literal sense?

Figure 1. Sample tasks used in the first training session.


The *second session* of the corpus training focused on the introduction to the COCA and its features. In this hands-on session, the participants were asked to consult the COCA to complete the tasks individually and discuss the data they obtained with the whole group.

Following the presentation of the specific features of the COCA, the researchers guided the PSTs to register to the corpus website and demonstrated several samples of queries by using different features of the COCA, such as list, collocation, word, and keyword in context (KWIC) searches. Figure 2 below displays a sample task used in the second session.

Word -> pretty



COCA - WORD QUERY



TASK 10

Better collocates

Look at the example sentences below. The underlined word in each sentence is an awkward collocate of the word in bold. Using COCA, find better collocates and revise the sentences.

1) I hope to succeed the **goal**.

→

2) There has been a hot **debate** over this issue.

→

3) He firmly **recommended** this place.

→

Figure 2. A sample task used in the second training session.

At the end of the second session, the participants were instructed to design one corpus-based material for teaching a commonly confused pair of lexical items such as *do* and *make*, *between* and *among*, etc., in pairs and groups of three, as shown in Figure 3 below.

TASK 11 *Let's design a DDL material!*

Design one corpus-based teaching material for one of the lexical item pairs below at any proficiency level that you want.

- **since & for**
- **between & among**
- **when & while**
- **affect & effect**
- **desert & dessert**
- **do & make**

1. *Language proficiency level*
2. *Aim of the activity* (Presenting new lexical items, practice, etc.)
3. *Target lexical items*
4. *Activity type* (gap-filling, matching, open-ended, multiple-choice, etc.)

please upload your material to the relevant folder (Let's Design Materials! Task folder) in METUClass

Figure 3. Designing a DDL material activity.

After 10-15 minutes, one PST from each group was instructed to upload the teaching material they had prepared to the allocated folder in METUClass (also known as ODTUClass), the learning management website of the Middle East Technical University. All participants could access these teaching materials for their future teaching experiences.

Tools & Instruments

The present study utilized multiple data collection tools and instruments for gathering data about the PSTs' perceptions regarding integrating corpora and corpus-based materials into language classrooms before and after the corpus training. The researchers consulted four experts in ELT and corpus pedagogy to establish the reliability and validity of the data collection instruments before their administration to the participants.

Background survey

The background survey was adapted from the corpus-based research literature (Bunting, 2013; Çalışkan, 2020; Geluso & Yamaguchi, 2014) to investigate the PSTs' general knowledge, experience, and perceptions towards using corpora. This survey provided demographic information about the participants and included questions regarding participants' background knowledge and awareness of corpora.

Semi-structured interviews before and after the corpus training

Online interviews were conducted twice with voluntary PSTs before and after the corpus training. Voluntary PSTs attended the interviews before (n=7) and after (n=10) the corpus training. Pre-training interviews included four questions regarding the participants' knowledge of corpora, DDL, and corpus tools. Post-training interviews comprised five questions on the training PSTs received and the integration of corpus into English language teaching. These interviews helped gather in-depth insight into PSTs' thoughts and emotions

regarding the training process. Voluntary PSTs attended the semi-structured interviews on the Zoom platform at the date and time they had scheduled via Doodle, and each interview lasted for 10-15 minutes. Based on the semi-structured interview protocol, all interviews were conducted in English and were audio-recorded and transcribed verbatim.

Reflective logs

PSTs were asked to write one reflective log at the end of the second training session to gain insight into their feelings about the training sessions, perceptions regarding integrating corpora into language classrooms, and remarks on their likelihood to consult corpora for their future teaching experiences. The participants were provided three guiding questions and sent their reflective logs via Google Forms in English.

Data Collection

The study lasted for a total of four consecutive weeks. The data collection procedure followed in the study is given in Table 2 below.

Table 2. Data Collection

Weeks	Data Collection Procedure
Week 1	Informing the PSTs about the study PST's signing the consent forms Administration of the background survey Pre-training semi-structured interviews
Week 2	Training Session 1
Week 3	Training Session 2
Week 3	Collection of the reflective logs
Week 4	Post-training semi-structured interviews

Data Analysis

The quantitative data gathered from the background survey to answer the first research questions were analyzed by descriptive statistics, and the qualitative data gathered through the interviews and reflective logs were analyzed by employing the principles of qualitative analysis for grounded theory (Smith, 2008) using the Constant Comparison Method (Boeije, 2002). Accordingly, the data were divided into small segments of single words/phrases (codes) that reflected the participants' perceptions regarding using corpora in language classrooms. Following the principles of the CCM, after listing all codes, the data analysis was carried out in a cyclical process of comparing codes to form sub-categories. These sub-categories were constantly compared and contrasted until they were delineated into the main categories. As a result, the categories emerged from the qualitative data rather than starting with preconceived categories (Charmaz, 2006). In order to ensure the reliability of the qualitative analysis, two raters separately analyzed the qualitative data by coding, categorizing, delineating themes, and connecting them. Any disagreements between the two raters regarding coding and categorizing were resolved by reviewing and negotiating for finalizing the codes for the emerging sub-categories and main categories. The raters employed Tawney and Gast's formula (1984) for measuring the percentage of inter-rater reliability, and it was found to be 0.86, indicating a high level of agreement between the raters.

Findings

Turkish Pre-Service EFL Teachers' Corpus Familiarity

A total of 63 PSTs out of 108 answered the background questionnaire. Descriptive statistics regarding the background survey revealed that most PSTs had no or little familiarity and experience using corpora and corpus-related tools or materials in English language teaching. Moreover, the participants lacked the knowledge of different usages of corpora in language classrooms. Table 3 below shows PSTs' knowledge of corpus and its use in L2 Classrooms.

Table 3. The PSTs' Knowledge about Different Usages of Corpora in L2 Classrooms

<i>Degree of knowledge</i>	Directly with students		Creating corpus-based materials		Creating language test items		Checking students' language use		Building language knowledge	
	<i>N*</i>	<i>%</i>	<i>N*</i>	<i>%</i>	<i>N*</i>	<i>%</i>	<i>N*</i>	<i>%</i>	<i>N*</i>	<i>%</i>
No knowledge	56	88.88	51	80.95	53	84.12	50	79.36	48	76.19
A little knowledge	4	6.34	9	14.28	7	11.11	8	12.69	10	15.87
Some knowledge	3	4.76	3	4.76	3	4.76	3	4.76	3	4.76
Extensive knowledge	-	-	-	-	-	-	2	3.17	1	1.58
Expert knowledge	-	-	-	-	-	-	-	-	1	1.58
Total	63	100	63	100	63	100	63	100	63	100

N*: Number of the participants

As visible in Table 3 below, the majority of the participants had no knowledge about using corpus tools for language teaching purposes (n=56), and they were not familiar with how to create corpus-based language teaching materials (n=51) or design language test items (n=53) as well. Moreover, the PSTs, in general, had no knowledge of using corpus tools (e.g., online corpora, DDL materials) to check students' language use (n=50).

Most participants did not use corpus for their language knowledge (n=48), and only a few (n=10) consulted it for language improvement. The participants did not have extensive or expert knowledge of the different uses of corpora in the language classroom. As well as their general knowledge and experience of corpus for specific purposes for language teaching, the PSTs were unfamiliar with the available corpus software (e.g., AntConc, SketchEngine) and online concordancing tools (e.g., BNC, COCA, NOW) prior to the study. In addition to PSTs' general knowledge and familiarity with corpus, they were also asked about their degree of knowledge of various corpus software and tools. Table 4 below shows their knowledge of such online corpus tools.

Table 4. The PSTs' Knowledge of Corpus Software and Online Corpus Tools

<i>Degree of knowledge</i>	Corpus software (e.g., AntConc, SketchEngine)		Online corpus tools (e.g., BNC, COCA)	
	<i>N*</i>	<i>%</i>	<i>N*</i>	<i>%</i>
No knowledge	59	93.65	53	84.12
A little knowledge	3	4.76	9	14.28
Some knowledge	1	1.58	1	1.58
Extensive knowledge	-	-	-	-
Expert knowledge	-	-	-	-
Total	63	100	63	100

N*: Number of the participants

According to Table 4, prior to the study, the PSTs, in general, had no knowledge of different corpus software (n=59) that would help them to conduct corpus-based analysis and concordancing or on available online corpus tools that would enable corpus inquiries on different linguistic levels (n=53). Even though a few students had little knowledge about online corpus tools (n=9), most had no extensive or expert knowledge of using any of the software programs or online tools. The following two excerpts below exemplify PSTs' prior knowledge of corpus software/tools:

We used it to check the frequency of some lexical items in an advanced grammar lesson, but it was just given as a "this exists, so you should be aware maybe" (PST3, Background Survey)

I have looked at the Türkçe Derlem Sözlüğü (Turkish Corpus Dictionary) for my linguistics class to be able to being familiar with a corpus dictionary (PST1, Background Survey)

As can be summarized from the results of the background survey, the majority of the participants had no or little contact with corpora prior to the training and had limited access to available corpus software and tools for language teaching. Students' familiarity with corpora is limited to instances of consulting corpus for their courses a few times. In addition to the descriptive findings in the background questionnaire, The PSTs were also interviewed before the training about what they think regarding the place of corpora in language teaching and the use and integration of corpus into English language teaching contexts. The purpose of these interviews was to have a deeper understanding of the lack of familiarity and experience with corpus-based language teaching and gather more information to identify the needs of the PSTs for the effective integration of corpus and DDL materials.

The qualitative data obtained from one open-ended item in the background survey and the pre-training interviews were analyzed to shed light on what the PSTs think about corpus use, corpus tools, and materials, and to what extent they were familiar with the use of corpus for language teaching purposes. Qualitative analysis resulted in 51 codes formed under five sub-categories and two main categories. Table 5 presents the qualitative findings related to PSTs' perceptions regarding integrating corpora into language classrooms before the training.

Table 5. Main and sub-categories related to the PSTs' perceptions before the training.

Main Categories	Sub-categories	N*
	<i>Corpora is:</i>	
Language Teaching	Helpful for language teaching	19
	Effective in language teaching	9
	Total	28
	<i>Corpora provides:</i>	
Usage	Access to real-life usage	12
	Access to a collection of words	6
	Dictionary-like use	5
	Total	23
	Main Total	51

N*: Number of codes

The main category regarding the benefits of integrating corpora and DDL tasks for **language teaching** included 28 codes and two sub-categories, namely indicating corpus was perceived as *helpful for language teaching* (19 codes) and *effective in language teaching* (9 codes). The PSTs only emphasized the benefits of corpora and DDL tasks to EFL language instruction based on their awareness and limited familiarity with corpus tools/online corpora, even though they had not had hands-on experience before the training. The PSTs claimed that corpora would help to provide language samples in contexts so that their learners would engage better with practical instruction and activities in the classroom:

I think it's helpful; it can be efficient in using you know to create real life contexts in lessons or in anywhere else in teaching. I think that it would be helpful for students to understand or internalize these subjects (PST3, Pre-training interview)

It would make teaching easier in my opinion, and it would enable better communication with students (PST1, Background Survey)

The second main category included 23 codes and three sub-categories in total. The PSTs emphasized the benefits of corpora related to its **usage** based on their limited contact with corpus tools before the training. The sub-categories that emerged were about how corpus would provide *access to real-life usage* (12 codes), *access to a collection of words* (6 codes), and *dictionary-like use* (5 codes). The PSTs' remarks were shaped around the idea that consulting corpora would offer authentic and real-life-like language samples and that it helps access a collection of words. That is, corpora would help them to see the daily usage of language and, in this way, can become a reference tool like a dictionary. The excerpts below exemplify the PSTs remarks about the authenticity of the corpora:

It would also allow them to have an idea about how to use the language in real life (PST 3, Pre-training interview)

As far as I remember, a corpus is the compilation of real-life words or phrases that is used in a community (PST36, Background Survey)

All in all, these findings show that even though the PSTs did not have extensive or expert knowledge on using and integrating corpus and DDL tasks into language teaching, they heard about corpus and were well aware that it would be effective for teaching. Thus, they were enthusiastic about learning more about corpus and hands-on usage of it for language teaching. Considering their lack of corpus literacy and the need to familiarize PSTs with using and integrating corpora into English language teaching, the corpus training was conducted with 108 pre-service EFL teachers. The following section reports the findings related to the participants' perceptions and reflections on receiving training about integrating corpora and DDL tasks into English language classrooms.

PSTs' Perceptions Regarding Corpus Use After the Training

The qualitative data from the reflective logs and the post-training interviews were analyzed to answer the second RQ. A total of 211 codes were identified, and these codes were formed under fourteen sub-categories and three main categories. The main categories that emerged were related to the benefits to teaching (113 codes), corpora use for future practices (51 codes), and drawbacks of corpora (47 codes). Table 6 below presents the three main categories and sub-categories of each main category that emerged.

Table 7. Main and sub-categories related to the PSTs' perceptions after the training.

Main Categories	Sub-categories	N*
Benefits to Teaching	<i>The use of corpora is helpful for:</i>	
	accessing contextualized language	25
	accessing authentic language samples	22
	teaching particular lexical items	20
	designing materials	19
	developing writing skills	14
	academic life	13
	Total	113
Corpora Use for Future Practices	<i>PSTs are willing to:</i>	
	use corpora in the future	30
	learn more about corpora	11
	participate in more training on corpora	10
	Total	51
Drawbacks of Corpora Use	<i>Corpora use is:</i>	
	complicated for corpus research	17
	not appropriate for low-level learners	9
	not user-friendly interface	9
	limited query limit	7
	time-consuming	5
Total	47	
	Main Total	211

N*: Number of codes

Benefits to Teaching

The main category of the benefits of the integration of corpora and DDL tasks to teaching included 113 codes and six sub-categories in total, revealing that the use of corpora was helpful for *accessing contextualized language* (25 codes), *accessing authentic language samples* (22 codes), *teaching particular lexical items* (20 codes), *designing materials* (19 codes) *developing writing skills* (14 codes) and *academic life* (13 codes). Following the training, the PSTs stated that they mostly benefited from it in terms of awareness of the numerous gains of consulting corpora, such as accessing contextualized and authentic language samples for teaching particular lexical items and designing teaching materials for teaching language skills and areas. The following excerpts show the PSTs' ideas on these issues.

I think using corpus-based materials is quite useful and authentic for students to provide them real and unique sources that are related to the usage of English in specific contexts (PST29, Reflective log)

Different levels need different things in the education system, maybe you have a language class, and you have to give importance on the vocabulary, so I think corpus is the best to prepare materials (PST1, Post-training interview)

In addition to providing authentic and contextualized language samples for vocabulary teaching and designing materials, the PSTs asserted that corpora would be useful for their academic life as a professional development opportunity. Moreover, using and consulting corpora was found useful, especially for their academic improvement, as shown in the following extracts:

I think that it is a great activity to do. It was amazingly helpful for English teacher candidates to learn how to use corpus effectively for our academic life and daily life purpose (PST33, Reflective Log)

In my future writing tasks, essays, and papers I think this training will help me a lot to improve the quality of my product (PST12, Reflective Log)

Before the training, PSTs knew the potential benefits of using corpora and DDL tasks. After the training, they had a chance to experience more beneficial uses of corpus-based language learning in teaching contexts and as a way to improve themselves academically. It was clear that the PSTs found the authentic and contextualized language samples in corpus data beneficial for language instruction and preparing teaching materials as well.

Corpora Use for Future Practices

The second main category regarding PSTs' perceptions was the overall evaluation of corpora use for future teaching practices since they were willing to *learn more about corpora* (11 codes), *participate in more training on corpora* (10 codes), and *use corpora in their future teaching experiences* (30 codes). The PSTs would like to continue in corpus-based language pedagogy when they become professionals. They were eager to implement corpora and DDL tasks in their actual teaching contexts when they become in-service teachers. The following excerpt illustrates this:

In my future teaching career, I will make use of the corpus tools as it contributes to both students' and the teacher's knowledge (PST40, Reflective Log)

It was revealed that most participants were excited about learning more about corpora use and implementing it in their future teaching careers and language classrooms. The remarks of the PSTs in their reflective logs and the interview showed that they enjoyed the training sessions and found the training content interesting and beneficial enough to feel the urge to engage more with what it provides.

Drawbacks of Corpora Use

The PSTs also stated some challenges they faced during the training related to using the corpus tool and corpus-based language teaching. Based on the drawbacks of using and integrating corpora and DDL tasks in teaching, the final main category included 47 codes and five sub-categories. The PSTs stated that corpora use is *complicated for corpus research* (17 codes), *not appropriate for low-level learners* (9 codes), *does not have a user-friendly interface* (9 codes), *has a limited query limit* (7 codes), and *time-consuming* (5 codes).

The drawbacks of corpora were shaped around the idea that it was hard to navigate the interface while searching for and analyzing corpus data due to the vast number of options to choose from, especially for elementary-level learners due to the language level used in the interface, according to the PSTs. The following sample excerpts show PST's concerns about the drawbacks they faced:

Search specifications are quite hard. It is a little complicated to use because there are lots of things you do not know where to look at. (PST4, Post-training interview)

I would like to use it myself; however, I do not think every level can use it or it can be widely used in every classroom. (PST1, Reflective log)

Actually I did not like the corpus website. It was hard to navigate in... it is not user friendly. it does not have an interface for newcomer users (PST9, Post-training interview)

The PSTs also claimed that corpus consultation could become time-consuming due to the multiple types and number of queries required for finding specific corpus data. Moreover, the search limit per user per day was a massive setback as the website did not allow participants to connect to the same network to carry out corpus queries after a specific number of log-ins. The premium version of the corpus was available through purchase, and such a cost was not affordable for the PSTs. The following excerpts highlight such drawbacks:

Another drawback could be the limits of work usage 50 queries as we also experienced it. (PST4, Post training interview)

I do not think this is easy or fast to use in classrooms. It is very complex and takes too much time. I do not think that I would need to use it. (PST 34, Reflective log)

All in all, the qualitative data analysis unveiled the perceptions of the pre-service EFL teachers regarding the benefits and drawbacks of integrating corpora and their overall evaluation and perspectives for future use. Despite a few challenges they experienced while having hands-on experience on the tasks provided by the researchers during the study, the PSTs' overall perceptions regarding and experience with the corpus use were mainly positive, and the number of benefits of corpora use in L2 classroom outnumbered the challenges stated by the pre-service EFL teachers. Most PSTs stated that they would consult corpora in their future

teaching experiences. Their statements showed they favored using corpus and preparing DDL tasks to provide authentic and contextualized language samples in their classrooms.

Discussion

This study's findings significantly revealed a need for corpus-related research in teacher education contexts. Prior to the training, the participants had almost no experience regarding the use of corpora for language teaching purposes, as echoed in several previous studies in the literature (Çalışkan & Kuru Gönen, 2018; Çalışkan, 2020; Özbay & Kayaoğlu, 2015; Zareva, 2017). This finding reflects that the introduction and integration of corpus-based language pedagogy have been neglected in many teacher education programs until recently. Even though the PSTs were aware of the potential uses and benefits of corpora, such awareness was quite limited as they had no chance to be involved in hands-on experiences of corpus consultation and designing language teaching materials.

Following the corpus training, including the theoretical aspects of corpora, presentation of a corpus tool, and practice activities to teach language skills and areas, corpora use was mainly reported to be fun, engaging, and beneficial by the PSTs, albeit with several drawbacks. The PSTs highlighted the benefit of consulting corpus to provide authentic language samples while preparing teaching materials in line with several previous studies' findings (Boontam & Phoocharoensil, 2018; Rasikawati, 2019). In addition to this, corpora use was found to be beneficial exclusively for teaching grammatical features of the language and vocabulary instruction in terms of teaching specific lexical units such as collocations and synonyms (Akkoyunlu & Kilimci, 2017; Lin & Lee, 2015; Soruç & Tekin, 2017; Yılmaz, 2017).

Consulting corpora has been found beneficial for preparing teaching materials and raising learners' language awareness in L2 classrooms (Akkoyunlu & Kilimci, 2017; Farr, 2008; Kızıl & Savran, 2018; Lewandowska, 2014; Mizumoto, Chujo & Yokota, 2016). As well as grammar and vocabulary instruction, the participants in the current study found the use of corpora helpful for improving their language skills and claimed that corpora use was likely to be highly beneficial for their academic improvement, a finding reported in many studies (Crosthwaite et al., 2021; Jamal et al., 2021; Yılmaz, 2017). Moreover, most PSTs questioned why corpora were not implemented in their courses as a part of their BA education. Their introduction to corpora opened new gates for them to improve themselves both as a learner and a teacher. That is, they realized that they would consult corpora for their studies in and outside the classroom (Bal-Gezegin et al., 2022; Varley, 2009). One PST mentioned in the post-training interview that she had even started using corpora for preparing materials and writing academic papers for her other courses after the first training session. Furthermore, having found the use of corpora and DDL tasks in language classrooms promising, the PSTs stated their motivation and willingness to use corpora for their future teaching experiences.

Despite many benefits of corpora claimed by the PSTs, some participants also expressed the challenges they experienced and the drawbacks of using corpora. Some PSTs claimed that advanced proficiency level users might take advantage of corpora more than low proficiency level learners as the interface and searching system of the COCA might not be appropriate for all users (Aşık, Vural & Akpınar, 2015; Crosthwaite et al., 2021). Corpus inquiry requires making multiple queries for preparing a material or consulting corpora for any reason; hence, the PSTs also mentioned the technical difficulties of the COCA they had encountered in terms

of its complicated interface, search limit, and the difficulty of finding the data through corpus queries aligned with the findings of previous research (Farr, 2008; Leńko-Szymańska, 2015; Lin & Lee, 2015; Poole, 2022; Zareva, 2017).

All in all, the present study's findings revealed various benefits and drawbacks of corpora use in the pre-service teacher education context. The findings highlight the promising use of corpora in language teaching contexts. The PSTs found corpora use and DDL tasks beneficial as they were willing to consult corpora, especially for vocabulary and grammar instruction, and to prepare authentic and contextualized teaching materials. Given the lack of courses on corpus-based language pedagogy in many pre-service teacher education contexts, this study offers various implications for teacher educators, administrators, and policymakers in higher education.

Conclusion

This study aimed to offer PSTs training on the theoretical aspects of corpora, provide hands-on experience in using a corpus to prepare DDL tasks and explore their perceptions regarding integrating corpora and corpus-based materials into language classrooms. The findings revealed that the PSTs' perceptions of corpora use and DDL tasks prior to the study were limited to what they recalled from the few instances of a corpus mentioned in their linguistics courses.

Following the training sessions, the PSTs' perceptions regarding the implementation of corpora into language classrooms were predominantly based on the benefits and overall positive evaluation of corpora, as the majority of the participants benefitted from the access to contextualized and authentic language samples in the corpus and emphasized the effectiveness of corpora use for teaching lexical items and developing writing skills. Most participants wanted to learn more about corpora in their future professional teaching careers.

In light of the findings, this study presents several implications for teacher training programs and teachers. One implication is that the integration of corpora into foreign language teacher training programs can be promising for teacher candidates as there is a need for such programs to include technology-based approaches to keep up with the recent educational innovations and advancements such as the corpora. Following the hands-on practice, almost all PSTs enjoyed the corpus consultation and were willing to use it for future teaching experiences. Thus, either an elective or compulsory course based on the use of corpora can be implemented into teacher education programs for PSTs to better engage with what corpora provide and its possible uses in L2 classrooms.

Another implication highlights that incorporating corpora into field-specific courses outlined in teacher training programs, such as material development, methodology, skill teaching, and practicum courses, can be highly beneficial to improve PSTs' teaching abilities as current and future EFL teachers. As corpus-based materials provide contextualized and authentic language samples for teachers, integrating corpora into such field-specific courses might allow PSTs to consult corpora and prepare authentic materials for courses based on teaching pedagogy. Despite the challenges they experienced and stated, the PSTs claimed that they could benefit from corpora to design materials, improve their vocabulary, and study commonly confused words both as a learner and a teacher. Therefore, another implication of this study is that consulting corpora for skill teaching instruction in language classrooms

would benefit learners to engage in real-life language samples in authentic teaching materials and for teachers to refer to while preparing vocabulary teaching materials and error-correction tasks.

It is now well-known that corpus-based language pedagogy is not introduced to pre-and in-service teachers and learners in most language teaching contexts. Even though this study offers promising findings to cater to such a need, the findings and implications are limited to the specific context of the present study. One limitation of the study is that the two-week instructional period of the corpus training was not sufficient to implement the idea of corpus integration into language teaching. This study was the first step in designing more comprehensive and systematic training for PST education. There is also the need to design a training framework for corpus integration for pre-and in-service teachers. Another limitation is that only one corpus tool was utilized in this study. Future studies can benefit other corpus tools and online corpora to provide greater insight into the integration of corpora in L2 classrooms and the perceptions of teachers and learners. The qualitative research undertaken herein offers valuable insights. It is recommended that forthcoming studies consider the incorporation of quantitative data collection techniques to augment the comprehensiveness of their findings. Thus, conducting further studies with more participants in a longitudinal fashion is essential, offering detailed instruction on corpus use and integration.

References

- Akkoyunlu, A. N., & Kilimci, A. (2017). Application of Corpus to Translation Teaching: Practice and Perceptions. *International Online Journal of Education and Teaching*, 4, 369-396.
- Alsolami, T., & Alharbi, A. (2020). Saudi EFL learners' Perceptions of The Use of Corpora in Academic Writing Teaching. *Studies in English Language and Teaching*, 8, 94.
- Aşık, A., Vural, A. S., & Akpınar, K. D. (2015). Lexical awareness and development through data-driven learning: Attitudes and beliefs of EFL learners. *Journal of Education and Training Studies*, 4(3), 87-96.
- Bal-Gezegin, B., Akbaş, E., & Başal, A. (2022). "Corpus made my job easier": Preservice language teachers' corrective feedback practices in writing with corpus consultation. *New Directions in Technology for Writing Instruction*, 279-295.
- Baker, P., Hardie, A., & McEnery, T. (2006). *A Glossary of Corpus Linguistics*. Edinburgh University Press.
- Bennett, G. R. (2010). *Using corpora in the language learning classroom: Corpus linguistics for teachers*. Michigan ELT.
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics - investigating language structure and use*. Cambridge University.
- Boontam, P., & Phoocharoensil, S. (2018). Effectiveness of english preposition learning through data-driven learning (DDL). *3L, Language, Linguistics, Literature*, 24(3).
- Boeije, H. (2002). A purposeful approach to the constant comparative method in the analysis of qualitative interviews. *Quality & Quantity*, 36, 391-409.
- Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning*, 67(2), 348-393.

- Bunting, J. D. (2013). *An investigation of language teachers' explorations of the use of corpus tools in the English for Academic Purposes (EAP) class*. [Unpublished Doctoral Dissertation]. Georgia State University, Atlanta.
- Charmaz, K. (2006). *Constructing Grounded Theory: A Practical Guide through Qualitative Analysis*. London: Sage Publications.
- Chen, M., Flowerdew, J., & Anthony, L. (2019). Introducing in-service English language teachers to data-driven learning for academic writing. *System*, 87, 102148.
- Cobb, T., & Boulton, A. (2015). Classroom applications of corpus analysis. In D. Biber & R. Reppen (Eds.), *The Cambridge Handbook of English Corpus Linguistics*, pp. 478–497.
- Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. Pearson Education.
- Crosthwaite, P., Luciana, & Schweinberger, M. (2021). Voices from the periphery: Perceptions of Indonesian primary vs secondary pre-service teacher trainees about corpora and data-driven learning in the L2 English classroom. *Applied Corpus Linguistics*, 1(1), 100003.
- Çalışkan, G. & Kuru Gönen, S. İ. (2018). Training teachers on corpus-based language pedagogy: Perceptions on using concordance lines in vocabulary instruction. *Journal of Language and Linguistic Studies*, 14(4), 190-210.
- Çalışkan, G. (2020). *An investigation into efl university instructors' perspective on the use of corpora in language teaching* [Unpublished master's thesis]. Anadolu University, Eskisehir, Turkey).
- 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.
- Forti, L. (2019). Learner attitudes towards data-driven learning: investigating the effect of teaching contexts. In F. Meunier, J. Van de Vyver, L. Bradley & S. Thouësny (Eds), *CALL and complexity - short papers from EUROCALL 2019* (pp. 137–143).
- Geluso, J., & Yamaguchi, A. (2014). Discovering formulaic language through data-driven learning: Student attitudes and efficacy. *ReCALL*, 26(2), 225–242.
- Gilquin, G., & Granger, S. (2010). How can data-driven learning be used in language teaching? *The Routledge Handbook of Corpus Linguistics*, 359-370.
- Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- Jamal, J., Shafqat, A., & Afzal, E. (2021). Teachers' perceptions of incorporation of corpus-based approach in English language teaching classrooms in Karachi, Pakistan. *Liberal Arts and Social Sciences International Journal (LASSIJ)*.
- Johns, T. (1991). From print out to handout: Grammar and vocabulary teaching in the context of data-driven learning. In T. Johns & P. King (Eds.), *ELR Journal 4: Classroom Concordancing* (pp. 27–46). University of Birmingham.
- Kızıl, A. Ş., & Savran, Z. (2018). The integration of corpora into EFL Speaking Instruction: A Study of Learner Perceptions. *International Online Journal of Education and Teaching*, 5, 376-389.
- Leńko-Szymańska, A. (2015). A teacher-training course on the use of corpora in language education: Perspectives of the students. In A. Turula, B. Mikolajewska & D. Stanulewicz (Eds.). *Insights Into Technology Enhanced Language Pedagogy* (pp. 129–144). Peter Lang GmbH, Internationaler Verlag der Wissenschaften.

- Leńko-Szymańska, A. (2017). Training teachers in data-driven learning: Tackling the challenge. *Language Learning & Technology*, 21(3), 217–241.
- Lewandowska, A.A. (2014). Using corpus-based classroom activities to enhance learner autonomy. *Koninskie Studia Jezykowe*, 3(2), 237-255.
- Lin, M. H., & Lee, J. (2015). Data-driven learning: Changing the teaching of grammar in EFL classes. *ELT Journal*, 69(3), 264-274.
- Liu, D., & Jiang, P. (2009). Using a corpus-based Lexicogrammatical approach to grammar instruction in EFL and ESL contexts. *The Modern Language Journal*, 93(1), 61-78.
- Mizumoto, A., & Chujo, K. (2015). A meta-analysis of data-driven learning approach in the Japanese EFL classroom. *English Corpus Studies*, 22, 1-18.
- Mizumoto, A., Chujo, K., & Yokota, K. (2016). Development of a scale to measure learners' perceived preferences and benefits of data-driven learning. *ReCALL*, 28(2), 227–246.
- O'Keeffe, A., McCarthy, M. & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. New York: Cambridge.
- Özbay, A., & Kayaoğlu, M. N. (2015). EFL teacher's reflections towards the use of computerized corpora as a teaching tool in their classrooms. *Cukurova University Faculty of Education Journal*, 44(1) 85-104.
- 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.
- Rasikawati, I. (2019). Corpus-based data-driven learning to augment L2 students' vocabulary repertoire. *International Dialogues on Education Journal*, 6(2), 83-98.
- Römer, U. (2010). Using general and specialized corpora in English language teaching: Past, present and future. In M. C. Campoy-Cubillo, B. Bellés For- tuño, & M. L. Gea-Valor (Eds.), *Corpus-based approaches to English language teaching* (pp. 18–38). London, England: Continuum.
- Smith, J. A. (2008). *Qualitative Psychology: A Practical Guide to Research Methods*. London: Sage Publications.
- Soruç, A., & Tekin, B. (2017). Vocabulary Learning through Data-Driven Learning in an English as a Second Language Setting. *Kuram Ve Uygulamada Egitim Bilimleri*, 17, 1811.
- Tawney, W.J. & Gast, L.D. (1984). *Single subject research in special education*. Columbus: Merrill.
- Varley, S. (2009). I'll just look that up in the concordancer: integrating corpus consultation into the language learning environment. *Computer Assisted Language Learning*, 22(2), 133–152.
- Yılmaz, M. (2017). The Effect of Data-driven Learning on EFL Students' Acquisition. *Eurasian Journal of Applied Linguistics*, 3(2), 75–88.
- Yoon, H., & Hirvela, A. (2004). ESL student attitudes toward corpus use in L2 writing. *Journal of Second Language Writing*, 13(4), 257-283.
- Zareva, A. (2017). Incorporating corpus literacy skills into TESOL teacher training. *ELT Journal*, 71(1), 69–79.