

CORPUS USED AS A DATA-DRIVEN LEARNING TOOL IN L2 ACADEMIC WRITING: EVIDENCE FROM TURKISH CONTEXTS

Gamze Emir^a, Gonca Yangın-Ekşi^b
(^agamzeemir@gazi.edu.tr; ^byonca@gazi.edu.tr)

^{a,b}Gazi University
Yenimahalle, Ankara, Turkey 06560

Abstract: This study investigated the effectiveness of using corpus as a data-driven learning (DDL) tool to enhance the academic writing skills of Turkish EFL learners. The study also explored learners' views of the potential use of corpus in L2 academic writing. To achieve these objectives, a mixed-method sequential explanatory design was employed, involving freshman student teachers enrolled in the Department of English Language Teaching at a state university in Turkey. The participants completed four argumentative essay writing tasks. Two tasks employed conventional techniques for error correction, while the other two utilized corpus as a reference tool for error correction. The latter two tasks were complemented by corpus training for the participants. The results indicated that using corpus as a DDL tool had a significant impact on the academic writing skills of Turkish EFL learners, with notable improvements observed in both grammar and vocabulary use. Participants also expressed positive feedback on the use of corpus as a DDL tool in enhancing their L2 academic writing.

Keywords: corpus linguistics, data-driven learning, L2 academic writing, error correction

DOI: <http://dx.doi.org/10.15639/teflinjournal.v34i2/209-225>

Academic writing in a second language (L2) is commonly regarded as a difficult skill to master, particularly for non-native English speakers who learn English in a foreign language learning context. Scholars have noted that L2 academic writing poses various sociocultural and linguistic difficulties for learners, involving issues such as word choice, grammatical accuracy, sentence structure, and discourse organization, among others (Cangır, 2023), leading to an extensive investigation of effective ways to improve L2 learners' academic writing. To address this issue, corpora and concordancers have gained widespread recognition as essential resources for effective L2 writing, with corpora playing a crucial role in providing authentic and diverse language samples for learners to analyze and learn (Boulton & Vyatkina, 2021; Römer, 2011). Despite the potential benefits of corpus approaches for L2 academic writing, scholars raise their concerns about a gap between advances in corpus studies and the practice of English language teaching, arguing that corpus applications have not yet been fully utilized in the language classroom (Chambers, 2019; Chen, et al., 2019; Hirata & Thompson, 2022; Pérez-Paredes, 2022).

Data-driven learning (DDL), a term coined by Johns (1991), aims to directly engage language learners with corpus data in order to eliminate intermediaries as much as possible

(Johns, 1991). By doing so, learners are empowered to study the target language and develop their own profiles of language use and meaning (Boulton & Cobb, 2017; Boulton & Pérez-Paredes, 2014). The multiple affordances of language corpora for DDL have emerged in various fields, including first and second language acquisition, skill development, translation, and materials development (Lenko-Szymańska & Boulton, 2015). The central idea of DDL is the hands-on use of authentic corpus data by advanced foreign or L2 language learners for self-directed language learning (Boulton, 2011). Regarding L2 acquisition, DDL creates an environment for students to become active language learners by enabling them to query corpora, understand concordance and collocation information, and notice gaps in their linguistic knowledge in the L2 repertoire (Godwin-Jones, 2017; O’Keeffe, 2021; Schmidt, 2023). This approach presents an alternative to teacher-led, rule-based approaches to language pedagogy (Crosthwaite, 2017).

In other words, DDL has the potential to enhance the academic writing skills of L2 learners by promoting independent learning through corpus-based discovery (Flowerdew, 2015). This approach encourages self-directed, inductive learning where students take responsibility for their own learning (Aston, 2015; Godwin-Jones, 2017). O’Keeffe (2021) notes that accessing a large amount of authentic corpus data can facilitate noticing (Schmidt, 1990, 2001), a process in which learners raise their awareness of so-called fossilized errors and explore their interlanguage features to improve their writing skills.

Numerous studies have documented the use of in-class corpus-driven DDL to improve academic writing. To achieve effective direct use of corpus in L2 writing, several studies have utilized corpus as a reference tool to correct learners' errors and enhance their writing skills. Kennedy and Miceli (2016) identified key characteristics of using a native corpus to improve learners' writing skills and revealed that working on corpus data was useful in promoting learners' development of a lexis-oriented view of the target language rather than a grammar-oriented view. Focusing on improving learners' use of linking adverbials in their essays by consulting a native corpus as a reference tool, Larsen-Walker (2017) found that learners improved their semantically appropriate production of linking adverbials in their writings. Similarly, Cotos (2014) conducted an experimental study comparing a control group who was using L1 reference corpora only with an experimental group who was using both L1 corpora and an L2 learner corpus derived from their students' personal data. This study provided support for utilizing learner corpus data to assist in identifying adverbial errors and ultimately in enhancing writing skills. Rather than solely relying on corpus as a reference source, Crosthwaite (2017) examined postgraduate language learners' error correction in their L2 writings through corpus consultation in combination with teacher feedback. The study revealed a positive effect of corpus consultation as a DDL tool in which learners corrected their lexical and grammar errors such as collocations, word choice, and word form. In a more recent study, Sun and Hu (2020) conducted research aimed at improving learners' use of hedging in academic writing. Their findings showed that consulting corpus data had a positive impact on the appropriate use of hedges and the quality of learners' academic writing. Each of these quantitative studies also includes qualitative data, indicating that both teachers and learners positively received the use of corpora for in-class DDL.

Several studies have demonstrated the positive impact of data-driven learning (DDL) on enhancing the language skills of L2 learners of Turkish. These studies provide evidence that

DDL holds promise in improving various aspects of foreign language learning, including collocation teaching (Özbay & Özer, 2017), enhanced lexical awareness (Aşık et al., 2016), vocabulary acquisition (Soruç & Tekin, 2017), instruction of academic lexical bundles (Lay & Yavuz, 2020), and grammar learning (Özer & Özbay, 2022). Collectively, these studies not only emphasize the potential benefits of DDL in foreign language instruction compared to traditional methods but also highlight the gaps in the existing empirical literature. The studies point towards the promising direction of research that aims to bridge the divide between advancements in corpus studies and the practice of English language teaching. They suggest further research to successfully integrate corpus applications and the DDL approach into mainstream education in Turkey.

Drawing on a limited number of research studies conducted on DDL approaches to foreign language learning, particularly in the context of L2 academic writing in Turkey, this study investigates the impact of data-driven learning (DDL) on improving Turkish EFL learners' academic writing. More than a decade ago, Römer (2006) stated that "a lot still remains to be done before [...] we can say that corpora have actually arrived in language pedagogy" (p. 129), and scholars today express concerns about the gap between theory and practice, suggesting that corpus applications have yet to become mainstream in second/foreign language teaching and learning (Perez-Paredes, 2022). To address this issue, this study aims to examine the effectiveness of corpus use as a DDL tool for improving L2 academic writing, including content, organization, grammar, vocabulary, and mechanics. Additionally, the study seeks to explore Turkish EFL learners' views on the potential of corpus use in L2 academic writing. The following research questions are addressed in this study:

1. Does the use of corpus as a DDL tool have an impact on enhancing Turkish EFL learners' L2 academic writing as compared to conventional techniques in terms of content, organization, grammar, vocabulary, and mechanics?
2. Which types of error are corrected by Turkish EFL learners when using corpus as a DDL tool?
3. What are Turkish EFL learners' views on the potential use of corpus as a DDL tool in L2 academic writing?

METHOD

Design

This study employed a mixed-method sequential explanatory design to scrutinize Turkish EFL learners' corpus use in their L2 academic writing. A mixed-method design integrates both quantitative and qualitative data at a particular stage of the research to gain a better understanding of the phenomenon being studied (Creswell & Creswell, 2017). The mixed-method sequential explanatory design consists of two distinct phases, a quantitative procedure followed by a qualitative procedure in order to gain a general understanding of the studied phenomena through quantitative approaches as well as to explain or elaborate on the statistical results for a deeper understanding (Creswell & Creswell, 2017).

Based on this design, the study examined the effect of corpus use on Turkish EFL learners' academic writing skills by applying a pre-experimental design to compare corpus use with conventional techniques and conducting content analysis to explore which traits of L2 academic writing were improved by using corpus as a DDL tool. Lastly, a written opinion survey was conducted to assess learners' opinions about corpus use.

Participants

This study selected participants through convenience sampling, which involves selecting individuals who are readily available and meet specific criteria (Dörnyei, 2007). Sixty-one freshman student teachers studying at the Department of English Language Teaching at a state university in Turkey participated in the study. Out of the 61 participants, 49 were female and 12 were male. The ages of the participants ranged from 18 to 20, and their English proficiency levels were B2. As part of the study, all participants received training in the use of the Corpus for Contemporary American English (COCA) as a data-driven learning tool.

Data Collection Tools

The study utilized COCA as the DDL tool. It was selected for its availability and coverage of over 560 million words of authentic texts, which are divided into sub-sections. The concordance lines from COCA were also used for presentation and practice purposes during the training. To familiarize learners with corpus use, corpus-based activities were developed.

A weighted rubric was developed to categorize learners' errors and grade their writings before and after the intervention, with a focus on five dimensions: content, organization, vocabulary, grammar, and mechanics. Weighted rubrics consider every criterion, but they assign more weight to specific criteria based on the teaching focus, the key components of the standard, and the timing of assessment (Burke, 2009). The 4-point weighted rubric included five dimensions: content, organization, vocabulary, grammar, and mechanics.

At the end of the study, a written opinion survey was developed to explore participants' views on the potential of corpus use as a DDL tool in writing. It was administered to all 61 participants, with a response rate of 100%. The survey consisted of seven open-ended questions to explore the role of the target corpus tool in the writing process. Opinion surveys are commonly conducted to determine the opinions of the target population on a specific situation or subject (Balci, 2007).

Experiment

In this experiment, learners were trained to write argumentative essays and then assigned four writing tasks. The first two tasks were conventional-based, where they wrote an argumentative essay on a given topic and revised and edited their own errors and their peers' errors using traditional techniques like consulting dictionaries and grammar books.

After completing the conventional writing tasks, learners received two weeks of corpus training. In the first part, they learned about corpora and how to search for the frequency of words, vocabulary items, collocations, prefixes, suffixes, synonyms of words, grammar structures, and syntactic and semantic patterns, as well as how to search across sub-sections of

the target corpora. They also practiced corpus searches and participated in some corpus-based activities that were prepared for the current study's purposes. Finally, they were assigned two error correction activities through corpus use as a DDL tool to correct errors utilizing COCA.

Following the corpus training, learners were assigned two additional corpus-based writing tasks. In the third and fourth writing tasks, they wrote an argumentative essay and then revised both their own writing and their peers' writing through COCA. Once they had completed their writing tasks, a written opinion survey was administered to participants to obtain their views on the potential use of corpus in L2 academic writing.

Data Analysis

In order to assess the effectiveness of corpus use as a data-driven learning (DDL) tool for enhancing Turkish EFL learners' L2 academic writing, a weighted rubric was used to score the first and final drafts of their writings. To ensure the reliability of the scores, two raters independently scored the writing, and the final score was determined as the average of the two raters' scores. The conventional-based and corpus-based written samples were compared using inferential statistics to reveal the efficacy of corpus-based applications. A normality test was initially conducted using SPSS (version 22.0), and since the sample size ($N > 50$) was greater than 50, the Kolmogorov-Smirnov test was used. The test was statistically significant, with a p -value less than .005, indicating that the normality assumption was not met. Therefore, non-parametric tests were used (Cokluk et al., 2010). Specifically, the Wilcoxon signed-rank test from non-parametric tests was employed to compare conventional-based writings with corpus-based writings.

To identify the error corrections made by the learners, the researchers collected a copy of their first drafts and asked them to correct their errors using either conventional techniques or a corpus used as a DDL tool before creating the second drafts. The first draft was analyzed for errors, and the researchers checked whether they were corrected in the second draft to calculate the learners' corrections. Qualitative content analysis was employed to categorize the learners' errors and corrections, which helped determine the types of errors the learners made, the types of errors they corrected, and the errors they did not correct.

To investigate Turkish EFL learners' perceptions on the potential of corpus use as a DDL tool, qualitative content analysis was employed to the data obtained from written opinion surveys. Qualitative content analysis was conducted to reveal meaning behind the data through seeking coding units and patterns (Hoffman et al., 2012). The researcher coded and categorized the learners' responses based on relevance. Approximately 15% of the data were analyzed by another expert to ensure the reliability of error categorization. Miles and Huberman's (1994) formula was used to compute the percentage of reliability of error categorization, which showed that 90% of the errors were reliable.

FINDINGS AND DISCUSSION

Findings

The present study aimed to investigate the effectiveness of using corpora as a DDL tool to enhance the academic writing skills of Turkish EFL learners. First, the study sought to address the research question of whether there was a significant difference between the use of corpus-based applications and conventional techniques in improving L2 academic writing. The study compared the writing scores of learners who utilized conventional techniques to those who used corpora as a DDL tool. The Wilcoxon signed-rank test was then employed to analyze the writing scores, and the findings are presented in Table 1.

Table 1. Findings of the Wilcoxon Signed-Rank Test in Self-correction Tasks

Test	Measurement	N	Mean Rank	Sum of Ranks	Z	P
Content	Negative ranks	5	14,00	70,00	-3,547	,000
	Positive ranks	24	15,21	365,00		
	Ties	32				
Organization	Negative ranks	5	9,50	47,50	-2,748	,006
	Positive ranks	17	12,09	205,50		
	Ties	39				
Vocabulary	Negative ranks	5	18,00	90,00	-4,323	,000
	Positive ranks	31	18,58	576,00		
	Ties	25				
Grammar	Negative ranks	8	15,00	120,00	-2,611	,009
	Positive ranks	22	15,68	345,00		
	Ties	31				
Mechanics	Negative ranks	5	12,00	60,00	-2,711	,007
	Positive ranks	18	12,00	216,00		
	Ties	38				
Total	Negative ranks	7	20,71	145,00	-4,966	,000
	Positive ranks	45	27,40	1233,00		
	Ties	9				

**Based on negative ranks*

The Wilcoxon Signed-Rank test was conducted to compare the conventional-based and corpus-based written samples in self-correction tasks. The written samples were evaluated in terms of content, organization, vocabulary, grammar, mechanics, and total score. Each aspect was scored based on negative ranks, with higher scores indicating better writing performance. The findings showed a significant improvement in the traits of Content ($Z = -3,547$; $p < .05$), Organization ($Z = -2,748$; $p < .05$), Vocabulary ($Z = -4,323$; $p < .05$), Grammar ($Z = -2,611$; $p < .05$), and Mechanics ($Z = -2,711$; $p < .05$). Additionally, the total score of the corpus-based written samples was significantly higher ($Z = -4,966$; $p < .05$), highlighting the effectiveness of using corpus as a DDL tool to enhance L2 academic writing.

Table 2 presents the results of the Wilcoxon Signed-Rank test that compared the use of conventional techniques and corpus in peer-correction tasks across various aspects of L2 academic writing, including content, organization, vocabulary, grammar, mechanics, and total score. The negative ranks represent a decrease in scores while positive ranks indicate an increase in scores.

Table 2. Findings of the Wilcoxon Signed-Rank Test in Peer-correction Tasks

Test	Measurement	N	Mean Rank	Sum of Ranks	Z	P
Content	Negative ranks	8	10,50	84,00	-1,225	,221
	Positive ranks	13	11,31	147,00		
	Ties	40				
Organization	Negative ranks	10	10,50	105,00	-1,091	,275
	Positive ranks	13	13,15	171,00		
	Ties	38				
Vocabulary	Negative ranks	3	14,00	42,00	-4,487	,000
	Positive ranks	29	16,76	486,00		
	Ties	29				
Grammar	Negative ranks	3	13,00	39,00	-4,288	,000
	Positive ranks	27	15,78	426,00		
	Ties	31				
Mechanics	Negative ranks	5	8,00	40,00	-1,291	,197
	Positive ranks	10	8,00	80,00		
	Ties	46				
Total	Negative ranks	13	20,46	266,00	-4,000	,000
	Positive ranks	40	29,13	1165,00		
	Ties	8				

**Based on negative ranks*

The findings in Table 2 showed a significant improvement in Vocabulary ($Z=-4.487$; $p<.05$) and Grammar ($Z=-4.288$; $p<.05$), demonstrating that corpus-based applications had a positive effect on those two aspects of academic writing. However, no significant difference was found in Content ($Z=-1.225$; $p>.05$), Organization ($Z=-1.091$; $p>.05$), and Mechanics ($Z=-1.291$; $p>.05$). Nonetheless, the total score showed a significant difference between the conventional-based and corpus-based written samples ($Z=-4.000$; $p<.05$), indicating that corpus use as a DDL tool was effective in enhancing Turkish EFL learners' L2 academic writing.

In addition to evaluating the effectiveness of using corpus as a DDL tool, this study also analyzed Turkish EFL learners' corrections through corpus consultation in their writing. The content analysis aimed to identify the types of errors corrected by learners using corpus. To achieve this, the first and final drafts of the participants' conventional-based and corpus-based writing tasks were categorized based on errors, and the corrections made by learners were recorded. The findings of the content analysis are presented in Table 3.

Table 3. Participants' Corrections in Conventional- and Corpus-based Writing Tasks

Category	Thematic Category												Instances	
		Correction				False-repairs				No Correction				
		Conventional tasks		Corpus-based tasks		Conventional tasks		Corpus-based tasks		Conventional tasks		Corpus-based tasks		
	W1	W2	W3	W4	W1	W2	W3	W4	W1	W2	W3	W4		
Organization	Content	13	27	-	-	-	-	-	-	45	30	37	47	
	Expression	7	10	-	-	-	-	-	-	49	46	31	38	
	Title	1	1	-	-	-	-	-	-	-	-	-	-	
Grammar	Preposition	22	34	57	42	3	6	1	5	70	93	47	37	For-to, Of-to
	Tense	26	34	28	51	2	5	-	2	50	44	25	15	Forced-would force, Is effect
	Agreement	25	38	40	50	5	5	1	-	44	28	16	17	It destroy(s), That-disappear
	Sentence formation	9	15	33	39	2	2	-	-	75	65	27	27	People so indolent, There two options
	Article	21	11	15	24	4	-	2	2	37	39	46	31	a-an-the
	Pronoun	10	3	10	17	1	-	-	-	17	16	8	4	It-they, Them brain
	Clause	3	4	9	10	1	1	-	1	12	9	8	3	People which – who, where
	Infinitive-Gerund	2	6	14	14	-	-	-	-	7	10	1	3	Try to discovering, To knowing
	Active-Passive	2	2	7	4	-	1	-	-	7	3	3	10	You are driving by your ambitions

Category	Thematic Category													Instances
	Correction				False-repairs				No Correction					
	Conventio- nal tasks		Corpus- based tasks		Conventio- nal tasks		Corpus- based tasks		Conventio- nal tasks		Corpus- based tasks			
	W1	W2	W3	W4	W1	W2	W3	W4	W1	W2	W3	W4		
														All (were) killed by someone
	Modals	4	1	2	-	-	2	-	-	2	7	1	1	Could-had to, Should-must
	Comparative	2	1	-	3	-	-	-	-	4	2	1	1	Easiest-easier
	Superlative	2	-	-	-	-	-	-	-	-	-	-	-	This is (the first) step to..
Vocabulary	Verb	20	11	87	93	5	1	1	-	160	178	99	72	Sight-see, Arrive-achieve
	Word form	35	55	49	68	2	2	-	-	60	49	36	23	Childs-children, Intelligent-intelligenc e
	Noun	20	10	47	41	1	-	-	-	74	45	26	28	Words-opinions, Trouble-suffer
	Linking words	7	8	17	21	2	-	1	-	58	44	15	8	So – but, But-or
	Adverb	5	4	18	13	3	1	-	1	34	24	11	6	Beside-next to, Too-as well
	Adjective	4	8	12	33	-	3	-	1	36	47	14	10	Important, essential
Mechanics	Spelling	22	22	37	45	-	-	-	-	28	9	18	8	Thibk, Sociaty
	Capitalizati on	7	14	5	2	-	2	-	-	31	21	9	1	
	Punctuation	8	8	19	25	2	1	1	1	71	93	93	73	Coma, Dot

Table 3 displays the frequency of error correction using conventional techniques (W1, W2) and corpus-based tasks (W3, W4) in tasks involving Turkish EFL learners. The results indicated that learners made more corrections overall when using corpus-based tasks compared to conventional techniques. Specifically, learners' corrections through corpus consultation increased in all error categories, with the most significant increases observed in vocabulary and grammar errors.

In terms of vocabulary errors, learners' correction rates increased for all types, with the most prominent increase seen in verb errors when using corpus compared to conventional techniques. For grammar errors, learners' correction rates increased for preposition, tense, agreement, and sentence formation errors when using corpus, although they faced difficulties in correcting article errors. The most common errors in the category of mechanics were related to spelling and punctuation.

Finally, this study explored Turkish EFL learners' views on the potential of corpus use as a DDL tool in L2 academic writing. To explore their opinions, a written opinion survey was employed, and the findings of content analysis were presented in Table 4.

Table 4. Turkish EFL Learners' Views on the Potential of Corpus Use in L2 Academic Writing

Category	Thematic Category		Sample Key Terms
		<i>f</i>	
Sources of conventional techniques	Dictionaries/online dictionaries	49	Tureng, Oxford, Cambridge, Colins,
	Internet	36	Grammarly.com Google translate
	Books	28	Grammar books ELS
	Their existing knowledge	19	Myself
Corrections through corpus	Vocabulary	41	Meaning and form Vocabulary usage Academic vocabulary Synonyms
	Grammar	36	Grammar rules Grammar structures
Effectiveness	Effective and useful	56	Provide various samples Improve writing skills Easy and comprehensible
	Ineffective	5	Hard to use Complicated Time-consuming
Benefits	Assessing and improving their writing performance	52	Find and see my mistakes Correcting errors
	Introducing academic language	6	Academic vocabularies Academic discourse
	Making the writing process easier	3	Make easier to find and correct mistakes

Category	Thematic Category		Sample Key Terms
		<i>f</i>	
			Giving more accurate feedbacks
	Exposing to authentic language	2	Sentences from native speakers
	Gaining a new perspective to the writing process	1	
Willingness	Willing to use corpus	56	Effective, easy and useful Provides academic language Advising their peers
	Unwilling to use corpus	5	Complicated Time-consuming

Table 4 provides an overview of the Turkish EFL learners' views and opinions of the potential of corpus use as a DDL tool in L2 academic writing, addressing various aspects of corpus use. Participants reported using conventional techniques such as dictionaries and online resources to correct their own and their peers' errors. Regarding corpus use, they stated that they mainly consulted the target corpus tool to correct vocabulary and grammar errors, and almost all found corpus to be an effective and useful tool in improving their writing skills. Only a few participants mentioned that the corpus was ineffective as a reference tool due to the complex interface of COCA. They noted that COCA was not user-friendly and that searching for errors was a time-consuming process.

The participants also reported numerous benefits of corpus use as a DDL tool. They mentioned that exposure to authentic language, achieved by observing real-world language examples, allowed them to grasp the nuances of natural language use and incorporate more authentic expressions into their writing. This exposure also led to an increased awareness of their interlanguage features. Through the analysis of language patterns, participants gained insights into their own language use, which enabled them to identify areas for improvement and refine their writing style. Moreover, the DDL approach rendered the writing process more manageable for participants. It facilitated the identification and correction of errors while also enabling them to obtain more precise feedback. Furthermore, participants noted the introduction of academic language as a valuable outcome of corpus use. They encountered academic vocabularies and discourse structures, equipping them with essential tools for expressing ideas in scholarly contexts. Lastly, a particularly intriguing observation was the new perspective gained on the writing process, which implied that participants experienced a transformative shift in their approach to academic writing. This shift could involve understanding writing as a dynamic process, appreciating the role of context, and adopting a more analytical stance towards their own work. Perhaps owing to these mentioned benefits, most participants expressed willingness to utilize the corpus. The primary reason cited was its effectiveness and usefulness in error correction.

Discussion

This study has unveiled the positive impact of utilizing a corpus as a DDL tool in enhancing the academic writings of Turkish EFL learners compared to traditional methods. The Wilcoxon-Signed rank test showed a significant improvement in the learners' L2 academic writing when using corpus as a DDL tool. The content analysis also demonstrated similar results, revealing an increase in learners' correction of their errors, particularly vocabulary and grammar errors, when using corpus to enhance their academic writing skills.

Consistent findings have been reported in the existing literature regarding the impact of utilizing corpora as DDL tools to enhance language learners' writing skills. Luo (2016) and Feng (2014) observed that consulting corpora improved the writing accuracy and fluency of EFL/ESL learners. Similarly, Cotos (2014) highlighted notable progress in the writing skills of learners exposed to corpus-based data. Despite the observed increase in error correction, the content analysis of the corpus-based writings in Cotos' (2014) study revealed that the participants made no corrections to almost half of their errors. This aligns with the findings of the current study suggesting that many errors in corpus-based writing remain uncorrected. This result could imply that learners may require input processing strategies (Van Patten, 2004) even when utilizing inductive noticing strategies to analyze concordance output.

This study revealed that using corpus as a DDL tool improved grammar and vocabulary compared to conventional techniques. These findings are consistent with the existing literature, which shows that corpus use can have a positive effect on language learners' grammar and vocabulary. Feng (2014) reported an increase in learners' vocabulary use through corpus consultation in writing processes. Similarly, Crosthwaite (2017) found that learners mostly consulted corpus to correct vocabulary errors. Özer and Özbay (2022) also showed the effectiveness of corpus consultation in improving learners' grammar use, and their study suggested that corpus use as a DDL tool was effective in improving learners' grammar.

This study showed that corpus use did not have an impact on improving learners' content, organization, and mechanics. The existing literature, on the other hand, provided evidence that corpora can be an effective reference tool for improving content and organization (Birhan et al., 2021), and punctuation marks (Celik & Ekatmis, 2013). The conflicting results could be derived from differences in the study designs, participant backgrounds, or the specific instructional methods employed in each study. It's also possible that the impact of corpus use on different aspects of writing, such as content, organization, and mechanics, varies depending on the proficiency level of the learners, the complexity of the writing task, or the amount of guidance provided during the corpus consultation process (Boulton, 2009; Chambers & O'Sullivan, 2004; Ma & Maie, 2021).

This study found that the use of corpora enhanced learners' usage of prepositions, tense, agreement, and sentence formation. However, learners encountered difficulties in correcting their article errors while using the corpus as a DDL tool during their writing processes. These results are in line with previous research in the literature. For instance, Crosthwaite (2017) reported similar findings in his study on DDL-mediated error correction across learners' written samples, demonstrating that learners successfully corrected grammar errors, including morphosyntactic errors such as tense, number, and agreement. Larsen-Walker (2017)

investigated the effect of DDL on students' ability to use linking adverbials correctly in their persuasive essays and found that corpus use as a DDL tool improved learners' use of linking adverbials in their writing.

The present study did not uncover any improvement in learners' writing with respect to the usage of articles, passives, quantifiers, and demonstrative determiners. In contrast, investigating the impact of corpus consultation as a DDL tool on language learners' grammar errors, Cowan et al. (2014) found that corpus use increased learners' awareness of grammatical errors in those four areas. This discrepancy might be due to the fact that the learners in the current study did not receive targeted instruction on these specific error types and were instead focused on correcting a range of language features during the error correction process.

The analysis of the participants' perspectives on using corpora as a DDL tool in their writing processes revealed favorable views toward corpora. Participants viewed corpora as practical and effective instruments for discerning their interlanguage features through comparisons between their native language and authentic data. Corpora can serve as an external feedback mechanism, aiding learners in recognizing distinctions between their language and the target language. They achieve this through the application of inductive learning mechanisms to deduce linguistic rules, the identification of patterns in the target language, and the comparison between their language and the target language to pinpoint disparities and inconsistencies (O'Sullivan, 2007). As a result, the utilization of corpora as a DDL tool could heighten their awareness of their interlanguage features, empowering them to identify and rectify linguistic issues (Flowerdew, 2015; Gilquin & Granger, 2010).

Nevertheless, a few participants reported challenges and found using the corpus to be time-consuming. These findings align with Aşık et al.'s (2016) study, which also noted that Turkish EFL learners encountered difficulties and time constraints when consulting corpora. The reported challenges in DDL may be attributed to potential confusion when interpreting concordance lines. Keywords are frequently presented within incomplete sentences, making it challenging to grasp the contexts of these lines (Hirata & Thompson, 2022). Additionally, EFL learners with limited exposure to the target language typically have more familiarity with traditional in-class activities or well-structured teacher-led tasks (Aşık et al., 2016; Hirata & Thompson, 2022). Although this study incorporated more learner-centered tasks aimed at promoting learner agency and self-regulation, as discussed by O'Keeffe (2021), some learners still found these tasks time-consuming and reported limited benefits from corpus use.

CONCLUSIONS

This study investigated the effectiveness of using a corpus as a tool for improving the L2 academic writing of Turkish EFL learners by comparing students' conventional-based and corpus-based written samples. This study also demonstrated participants' views towards the potential of using corpus as a tool for L2 academic writing. The study used the Wilcoxon signed-rank test as a non-parametric test to determine the efficacy of corpus use, and content analysis was employed to categorize learners' errors and corrections, indicating in which traits of L2 academic writing participants used corpus to correct their errors. Another content analysis was employed to demonstrate their views towards using a corpus as a DDL tool. The findings

revealed that the use of corpus had a significant positive impact on the L2 academic writing of Turkish EFL learners, especially in terms of increasing their vocabulary and grammar usage. Furthermore, the participants reported positive views on the usefulness of corpus use as a DDL tool in improving their L2 academic writing.

This study suggests certain pedagogical implications for error correction, and grammar and vocabulary teaching in light of the findings. Firstly, this study provides evidence that EFL learners can effectively consult corpora as a DDL tool to correct their errors and improve their writing skills. Therefore, it is recommended that corpora use and DDL activities be integrated into language learners' writing processes to enhance their language learning outcomes. Additionally, input processing and enhancement strategies are suggested when working on concordancers to improve EFL learners' correction rates. Second, this study provides evidence that integrating corpora improves learners' understanding of vocabulary and grammar within context. Therefore, the utilization of corpora and engagement in DDL activities could be recommended to facilitate vocabulary and grammar learning. Furthermore, this study reinforces the notion that integration of corpora and corpus-based activities can effectively promote learners' awareness of authentic language and its interlanguage features. Thus, it is recommended that material developers utilize corpora input to present authentic and meaningful data to language learners.

This study also suggests avenues for further research. First, this study is limited to including only an experimental group. A control group can be proposed to mitigate the learning effect. Also, this study has been carried out in a university setting. Further studies should be conducted to focus on lower-level language learners and to define language awareness. Further, corpus use was found to be ineffective by a few language learners in this study. Therefore, further research could investigate the effects of individual differences among language learners, such as learning styles, learning strategies, etc., to give more insights into the efficacy of corpus use on language teaching/learning. Lastly, this study, like most previous studies, focused on utilizing corpus as a reference tool in the language learning process. Further research can focus on corpora teaching to learners to get the maximum benefit of corpora in language learning/teaching.

NOTES

This paper is part of the first author's MA thesis under the supervision of the second author.

REFERENCES

- Aşık, A., Vural A. S., & Akpınar, K. D. (2016). Lexical awareness and development through data driven learning: Attitudes and beliefs of EFL learners. *Journal of Education and Training Studies*, 4(3), 67-96. <https://doi.org/10.11114/jets.v4i3.1223>
- Aston, G. (2015). Learning phraseology from speech corpora. In A. Lenko-Szymanska & A. Boulton (Eds.), *Multiple affordances of language corpora for data-driven learning* (pp. 65-84). John Benjamins.
- Balcı, A. (2007). *Sosyal bilimlerde araştırma: Yöntem teknik ve ilkeler* [Research in social sciences: Methods, techniques and principles]. Pegem Akademi Yayıncılık.

- Birhan, A. T., Teka, M., & Asrade, N. (2021). Effects of using corpus-based instructional mediation on EFL students' academic writing skills improvement. *Theory and Practice of Second Language Acquisition*, 7(2), 133-153. <https://doi.org/10.31261/TAPSLA.9993>
- Boulton, A. (2009). Testing the limits of data-driven learning: Language proficiency and training. *ReCALL*, 21(1), 37-54. <https://doi.org/10.1017/S0958344009000068>
- Boulton, A. (2011). Data-driven learning: The perpetual enigma. In S. Gozdz-Roszkowski (Eds.), *Explorations across languages and corpora* (pp. 563-580). Peter Lang.
- 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>
- Boulton, A., & Pérez-Paredes, P. (2014). ReCALL special issue: Researching uses of corpora for language teaching and learning editorial researching uses of corpora for language teaching and learning. *ReCALL*, 26(2), 121-127. <https://doi.org/10.1017/S0958344014000068>
- Boulton, A., & Vyatkina, N. (2021). Thirty years of data-driven learning: Taking stock and charting new directions over time. *Language Learning & Technology*, 25(3), 66-89. <https://doi.org/10125/73450>
- Burke, K. (2009). *How to assess authentic learning*. Sage Publications.
- Cangır, H. (2023). Using corpus tools for academic writing in EFL settings: A data-driven learning approach. In G. Yangın-Ekşi, S. Akayoğlu, & L. Anyango (Eds.), *New directions in technology for writing instruction* (pp. 171-191). Springer International Publishing.
- Celik, S., & Elkatmis, M. (2013). The effect of corpus assisted language teaching on the learners' proper use of punctuation marks. *Educational Sciences: Theory and Practice*, 13(2), 1090-1094. <https://files.eric.ed.gov/fulltext/EJ1017262.pdf>
- Chambers, A. (2019). Towards the corpus revolution? Bridging the research–practice gap. *Language Teaching*, 52(4), 460-475. <https://doi.org/10.1017/S0261444819000089>
- Chambers, A. & O'Sullivan, I. (2004). Corpus consultation and advanced learners writing skills in French. *ReCALL*, 16(1), 258-172. <https://doi.org/10.1017/S0958344004001211>
- Chen, M., Flowerdew, J., & Anthony, L. (2019). Introducing in-service English language teachers to data-driven learning for academic writing. *System*, 87, 102-148. <https://doi.org/10.1016/j.system.2019.102148>
- Cokluk, Ö., Şekercioğlu, G., & Büyüköztürk, Ş. (2010). *Multivariate statistics for the social sciences: SPSS and LISREL applications*. Pegem Akademi Yayıncılık.
- Cotos, E. (2014). Enhancing writing pedagogy with learner corpus data. *ReCALL*, 26(2), 202-224. <https://doi.org/10.1017/S0958344014000019>
- Cowan, R., Choo, J., & Lee, G. S. (2014). ICALL for improving Korean L2 writers' ability to edit grammatical errors. *Language Learning & Technology*, 18(3), 193-207. <http://llt.msu.edu/issues/october2014/cowanchoolee.pdf>
- Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage Publications.
- Crosthwaite, P. (2017). Retesting the limits of data-driven learning: Feedback and error correction. *Computer Assisted Language Learning*, 30(6), 447-473. <https://doi.org/10.1080/09588221.2017.1312462>
- Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford University Press.

- Feng, H. (2014). The use of corpus concordancing for second language learners' self error-correction. *Journal of Interactive Learning Research*, 25(1), 5-25. <https://www.learntechlib.org/primary/p/39379/>
- Flowerdew, L. (2015). Data-driven learning and language learning theories. In A. Lenko-Szymanska & A. Boulton (Eds.), *Multiple affordances of language corpora for data-driven learning* (pp. 15-36). John Benjamins.
- Gilquin, G., & Granger, S. (2010). How can data-driven learning to be used in language teaching. In A. O'Keeffe & 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. <http://llt.msu.edu/issues/october2017/emerging.pdf>
- Hirata, Y., & Thompson, P. (2022). Communicative data-driven learning: A two-year pilot study. *ELT Journal*, 76(3), 356-366. <https://doi.org/10.1093/elt/ccab066>
- Hoffman, J. W., Wilson, M., Martinez, R., & Sailors, M. (2011). Content analysis: The past, present, and future. In M. Mallette & N. Duke (Eds.), *Literacy research methodologies* (pp. 28-49). Guilford Press.
- Johns, T. (1991). Should you be persuaded: Two samples of data driven learning materials. In T. Johns & P. King (Eds.), *ELR journal vol. 4: Classroom concordancing* (pp. 1-16). Birmingham University Press.
- Kennedy, C., & Miceli, T. (2016). Cultivating effective corpus use by language learners. *Computer Assisted Language Learning*, 30(1), 91-114. <https://doi.org/10.1080/09588221.2016.1264427>
- Larsen-Walker, M. (2017). Can data driven learning address L2 writers' habitual errors with English linking adverbials? *System*, 69, 26-37. <https://doi.org/10.1016/j.system.2017.08.005>
- Lay, K. J., & Yavuz, M. A. (2020). Data-driven learning of academic lexical bundles below the C1 level. *Language Learning & Technology*, 24(3), 176-193. <https://doi.org/10.125/44741>
- Leńko-Szymańska, A., & Boulton, A. (2015). Introduction: Data-driven learning in language pedagogy. In A. Lenko-Szymanska & A. Boulton (Eds.), *Multiple affordances of language corpora for data-driven learning* (pp. 1-14). John Benjamins.
- Luo, Q. (2016). Application of data-driven learning in EFL writing instructions for engineering students. *World Transactions on Engineering and Technology Education*, 3, 238-244. [http://www.wiete.com.au/journals/WTE&TE/Pages/Vol.13,%20No.3%20\(2015\)/04-Luo-Q.pdf](http://www.wiete.com.au/journals/WTE&TE/Pages/Vol.13,%20No.3%20(2015)/04-Luo-Q.pdf)
- Ma, Q., & Mei, F. (2021). Review of corpus tools for vocabulary teaching and learning. *Journal of China Computer-Assisted Language Learning*, 1(1), 177-190. <https://doi.org/10.1515/jccall-2021-2008>
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Sage Publications.
- O'Keeffe, A. (2021). Data-driven learning—a call for a broader research gaze. *Language Teaching*, 54(2), 259-272. <https://doi.org/10.1017/S0261444820000245>

- O’Sullivan, Í. (2007). Enhancing a process-oriented approach to literacy and language learning: The role of corpus consultation literacy. *ReCALL*, 19(3), 269–286. <https://doi.org/10.1017/S095834400700033X>
- Özbay, A. Ş., & Özer, M. (2017). Effects of data-driven learning on collocational competence of EFL learners in a Turkish context. *The International Journal of Research in Teacher Education*, 8(3), 27-39. <http://ijrte.inased.org/makale/392>
- Özer, M., & Özbay, A. S. (2022). Exploring the data-driven approach to grammar instruction in the ELT context of Turkey. *Australian Journal of Applied Linguistics*, 5(2), 35-63. <https://doi.org/10.29140/ajal.v5n2.713>
- Pérez-Paredes, P. (2022). A systematic review of the uses and spread of corpora and data-driven learning in CALL research during 2011–2015. *Computer Assisted Language Learning*, 35(1-2), 36-61. <https://doi.org/10.1080/09588221.2019.1667832>
- Römer, U. (2006). Pedagogical applications of corpora: Some reflections on the current scope and a wish list for future developments. *Zeitschrift für Anglistik und Amerikanistik*, 54(2), 121-134. <https://doi.org/10.1515/zaa-2006-0204>
- Römer, U. (2011). Corpus research applications in second language teaching. *Annual Review of Applied Linguistics*, 31, 205-225. <https://doi.org/10.1017/S0267190511000055>
- Schmidt, N. (2023). Unpacking second language writing teacher knowledge through corpus-based pedagogy training. *ReCALL*, 35(1), 40-57. <https://doi.org/10.1017/S0958344022000106>
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129-158. <https://doi.org/10.1093/applin/11.2.129>
- Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3-32). Cambridge University Press.
- Soruç, A., & Tekin, B. (2017). Vocabulary learning through data-driven learning in an English as a second language setting. *Educational Sciences: Theory & Practice*, 17(6), 1811-1832. <https://jestp.com/article-detail/?id=400>
- Sun, X., & Hu, G. (2020). Direct and indirect data-driven learning: An experimental study of hedging in an EFL writing class. *Language Teaching Research*, 27(3), 660-688. <https://doi.org/10.1177/1362168820954459>
- Van Patten, B. (2004). Input processing in second language acquisition. In B. Van Patten (Ed.), *Processing instruction: Theory, research, and commentary* (pp. 5-31). Lawrence Erlbaum Associates.