Review

EFL Learners' Collocation Acquisition and Learning in Corpus-Based Instruction: A Systematic Review

Wei Sun [1] and Eunjeong Park [2],*

[1] School of English, Jilin International Studies University, Changchun 130117, China; zenithsw@sina.cn
[2] English Department, Sunchon National University, Suncheon-si 162742, Republic of Korea
* Correspondence: parkej@scnu.ac.kr

Abstract: This systematic review critically examines 32 previous empirical studies on corpus-based collocation acquisition and learning in EFL contexts, aiming to identify the research contexts, corpus types, instructional methods, and how corpus-based training affects EFL students' acquisition and learning of collocations. This review provides a comprehensive understanding of corpus-based collocation instruction and its effectiveness in enhancing learners' collocation acquisition and learning. It makes use of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) approach to achieve these objectives. The findings reveal a progressive growth in research and publications on corpus-based collocation acquisition and learning, indicating its increasing significance in language education. This review highlights the wide applicability of corpus-based approaches across different language learning contexts, particularly among English language learners in foreign language settings. It also underscores the global significance of corpus-based collocation instruction, with attention and research conducted in various countries, notably, in Asian regions. This review discusses the positive effects of corpus-based tools and data-driven learning on collocation acquisition and learning, demonstrating the improvement in learners' collocation knowledge, usage, and retention by comparing and summarizing research results. However, conflicting results regarding implicit and explicit instruction methods warrant further investigation. The collective findings of the reviewed studies contribute valuable insights into the effectiveness of different instructional methods and call for further research to optimize collocation instruction strategies with both theoretical and pedagogical implications.

Keywords: collocation acquisition and learning; corpus-based instruction; EFL learners; PRISMA; systematic review

1. Introduction

1.1. The Significance of Collocation in L2 Learning

Many scholars have emphasized the significance and worth of collocations in fostering L2 vocabulary growth and communicative proficiency [1–3], and they expressed their views on instructing collocations to non-native speakers, highlighting its advantages in providing pre-established language patterns and leading to enhanced performance. Hashemi [4] conducted a study that found both English majors and non-English majors among EFL college students, along with high school students and professors, exhibit a deficiency in collocational knowledge. This deficiency can be attributed to the neglect of collocation instruction in EFL classrooms.

In recent times, scholars and educators have emphasized the significance of collocations in the progress and instruction of language, especially when working with individuals who are not native speakers [5]. According to Leff's [6] research findings, collocation was considered more valuable than encyclopedic knowledge. Ellis [7] theorized that collocational knowledge forms the core of language proficiency. Nation [8] emphasized the

significance of using collocations correctly to achieve fluency, stating that "collocational knowledge is essential for all fluent and appropriate language use" (p. 132).

Collocation is an inherent linguistic pattern where words consistently appear together, forming interconnected units. This phenomenon is widespread across languages, and it is accurate to assert that collocations are an integral part of natural languages. In reality, "words seldom occur in isolation" (Wallace 1982, p. 30) [9]. Collocation also offers an effective strategy for memorizing new words. In support of this, Taylor, as cited by Nation (2004, p. 38) [10], outlines the rationale behind studying words in collocation, "words which are naturally associated in context are learnt more easily than those not so associated; vocabulary is best learned in context; context alone is insufficient without deliberate association ...". Moreover, research has shown that a lack of proficiency in collocations contributes to errors among EFL learners, impeding their fluency. In various instances, learners tend to apply English collocations to their native language equivalents when they are unsure about which words naturally co-occur in context. This highlights the importance of incorporating collocations into teaching materials and emphasizing their instruction and practice [9].

Hence, collocations hold significant importance and possess distinctive qualities, indicating that their usage plays a crucial role in enhancing a learner's language fluency and enabling them to approach native-like proficiency.

1.2. Challenges in Collocation Acquisition and Learning

Regrettably, collocations present significant difficulties for second language (L2) learners, who are frequently observed to improperly use, excessively utilize, or inadequately apply these word collocations [11,12]. Numerous studies have demonstrated the insufficient understanding of collocations among English as a Foreign Language (EFL) learners, and the findings from these studies have substantiated the ongoing debate surrounding collocations for language learners, particularly in EFL environments. In environments where English is learned as a foreign language (EFL), learners primarily come across isolated words rather than collocations. Consequently, encountering collocations in various contexts often proves to be a challenging task for them [13]. Clearly, a comprehensive grasp of collocations is essential for both expressing oneself in language and understanding its meaning [14], but based on the findings of numerous studies, learners encounter limitations in their ability to effectively use collocations. They face numerous challenges when it comes to producing and understanding collocations [15–17].

Hence, researchers have put forth several potential factors contributing to the challenges in acquiring collocations among ESL learners. As noted by Boers, Lindstromberg, and Eyckmans (2014) [18], one factor is the learners' tendency to overlook collocations due to lack of attention [19]. Semantic factors could also be at play. Numerous English verb–noun collocations (such as "make a mistake" or "have a nightmare") include what are known as light verbs (like "make" and "have"), which do not significantly contribute to the overall meaning of the collocation. As the noun carries the primary meaning, the verb often fails to capture learners' focus, potentially leading L2 learners to use an incorrect verb (such as "*do a mistake") [18]. Moreover, the learners' native language (L1) could potentially impede the acquisition of L2 collocations. Research indicates that ESL learners tend to process coherent L2 collocations with greater speed and accuracy than those that are incongruent [20,21].

1.3. Collocation Acquisition and Learning in the Corpus-Based Approach

A growing body of research has investigated the connection between teaching collocations and the process of acquiring and learning L2 collocations [22–24]. The findings among studies were inconsistent due to variables like intervention approaches, such as getting students to notice the target collocations using input enhancement [25], encouraging them to use corpus tools [26], or helping them memorize the target collocations by giving them tasks that were thought to improve retention [27]. Consequently, in the last decade, the

discourse surrounding the effectiveness of L2 collocation instruction has progressively transitioned toward evaluating the relative impact of different intervention approaches and addressing more specific factors that could influence effectiveness, including task types [27–29], mode of input [23,30], types of target collocations [31,32], and frequency of occurrence. Previous research has highlighted the challenges faced by language teachers in teaching multi-word verbs to their students, despite the availability of various learning materials such as textbooks [33], and specialized learner dictionaries [34,35]. According to Girgin [36], teaching these structures to students is extremely challenging for language teachers. The fact that these resources frequently present the concepts in a way that necessitates rote memorization and shows a lack of a systematic and effective strategy for learning them is one area of agreement [37].

Johns [38] popularized the use of corpora in language instruction and coined the term "data-driven learning" (DDL) to describe the corpus approach. The concordancer, a computer application used for searching and analyzing a certain word or phrase, is a vital tool for accessing a corpus. Corpus-based DDL has gained recognition as a potent instructional method for enhancing vocabulary knowledge [39] and language learning autonomy or independent learning [40]. The use of a corpus-based DDL method aids language learners in a number of ways [41]. First, by incorporating real language data, the corpus provides an authentic learning context, thus enhancing the authenticity of the learning experience. Second, an analysis of the corpus using concordance programs or concordancers equips L2 learners with valuable analytical skills. Third, corpus-based DDL facilitates the development of learners' awareness of the target language. Last, increased exposure to authentic texts through concordance, which can be accessed independently by learners, boosts motivation and promotes learner autonomy.

There are two types of pedagogical corpus applications: indirect applications and direct applications [42]. Other researchers use the similar terms hands-off and hands-on, respectively [43]. Recent studies have exhibited a heightened interest in the latter. The studies included in this review include both indirect applications and direct applications. The hands-on method gives students immediate computer access to a corpus with the goal of, for instance, learning vocabulary definitions. On the other hand, the hands-off method necessitates that students work with printed-out corpus materials created by their teacher. Using a computer-based corpus helps motivate students and provides them with a sense of independence and control over their education. However, not every classroom has a computer or access to the Internet [44]. Hence, printed materials can still be used to facilitate learning within a data-driven learning (DDL) classroom, particularly in settings where technological resources are limited. However, with the evolution of technology, the advancement of teaching facilities, and learners' easy access to/availability of mobile devices such as smartphones, iPads, and laptops, it is necessary to re-evaluate the effectiveness of hands-on and hands-off teaching approaches.

As a popular approach in the field of language instruction, corpus-based learning has gained significant attention in relation to collocation instruction. It is crucial to examine the main issues (e.g., research context, effects) of corpus-based instruction on collocation acquisition and learning.

2. The Aim of This Systematic Review

This study, as a systematic review, is dedicated to reviewing previous empirical studies on collocation acquisition and learning based on corpora (DDL). The purpose of this systematic review was to identify the research contexts in which corpus-based learning approaches were utilized, focusing on the corpora that to identify the main types of instructional methods that were utilized in studies on collocation instruction; and to assess the influence of corpus-based instruction on the acquisition of collocations by EFL learners. By systematically reviewing the existing literature, this study aimed to provide a comprehensive understanding of corpus-based collocation instruction and its effectiveness in enhancing EFL learners' collocation acquisition. Based on our research objectives,

this research synthesis on the corpus-based learning approach for EFL learners' English collocation acquisition aimed to answer four questions, which are listed in Table 1:

Table 1. Research Questions and Objectives.

RQs	Objective
RQ 1. What are the research contexts regarding the corpus-based learning approach for EFL learners' English collocation acquisition?	The objective of this question is to inquire about the research contexts related to the corpus-based learning approach for the acquisition of English collocations by EFL learners.
RQ 2. What type of corpus is utilized to improve ELF learner's collocation in previous research?	The objective of this question is to understand the corpus tools that were utilized in previous research to enhance English collocation acquisition for EFL learners.
RQ 3. What types of instructional methods are used in these studies?	The objective of this question is to identify and describe the main types of instructional methods used in studies related to collocation instruction.
RQ 4. What impact does corpus-based instruction have on the acquisition of collocations among EFL learners?	The objective of this question is to explore and understand the impacts of corpus-based instruction on the learning of collocations by EFL learners.

3. Methodology

3.1. The Framework of This Research Synthesis

Using a systematic review methodology, this study follows the guidelines specified in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework. The PRISMA framework was initially introduced by Moher et al. in 2009 [45] and was since revised in 2020, as highlighted by Page et al. in 2021 [46]. The PRISMA checklist comprises 27 items designed to enhance transparency and consistency in systematic reviews. Following a reporting guideline is essential to guarantee transparency, clarity, and thoroughness in systematic review articles. This study chooses to follow the PRISMA guideline because of its increasing acceptance among scholars engaged in systematic reviews, as evident from the numerous endorsements and citations it has garnered [46]. Additionally, methodology experts and journal editors regularly review and update this guideline [47], thus ensuring its relevance and rigor. The subsequent sections outline the application of the PRISMA framework in this study, as illustrated in Figure 1.

3.2. Phase 1: The Identification Phase

To comprehensively retrieve studies that are related to the corpus-based learning approach in EFL learners' English collocation acquisition, we searched several databases (Web of Science, Scopus, Wiley, Sage, ProQuest Dissertation and Theses, EBSCO, Taylor & Francis, and the Education Resources Information Center (ERIC)) and journals for publications in the last two decades (2003–2022). We chose this time frame due to the recent fast development trend in computer-aided education, and because 2003 was one year after the 12th Meeting on Computational Linguistics in the Netherlands (CLIN), where the theme focused on corpus-based acquisition of collocational prepositional phrases. Certain papers were also manually reviewed to identify studies that were not located using Google Scholar searches.
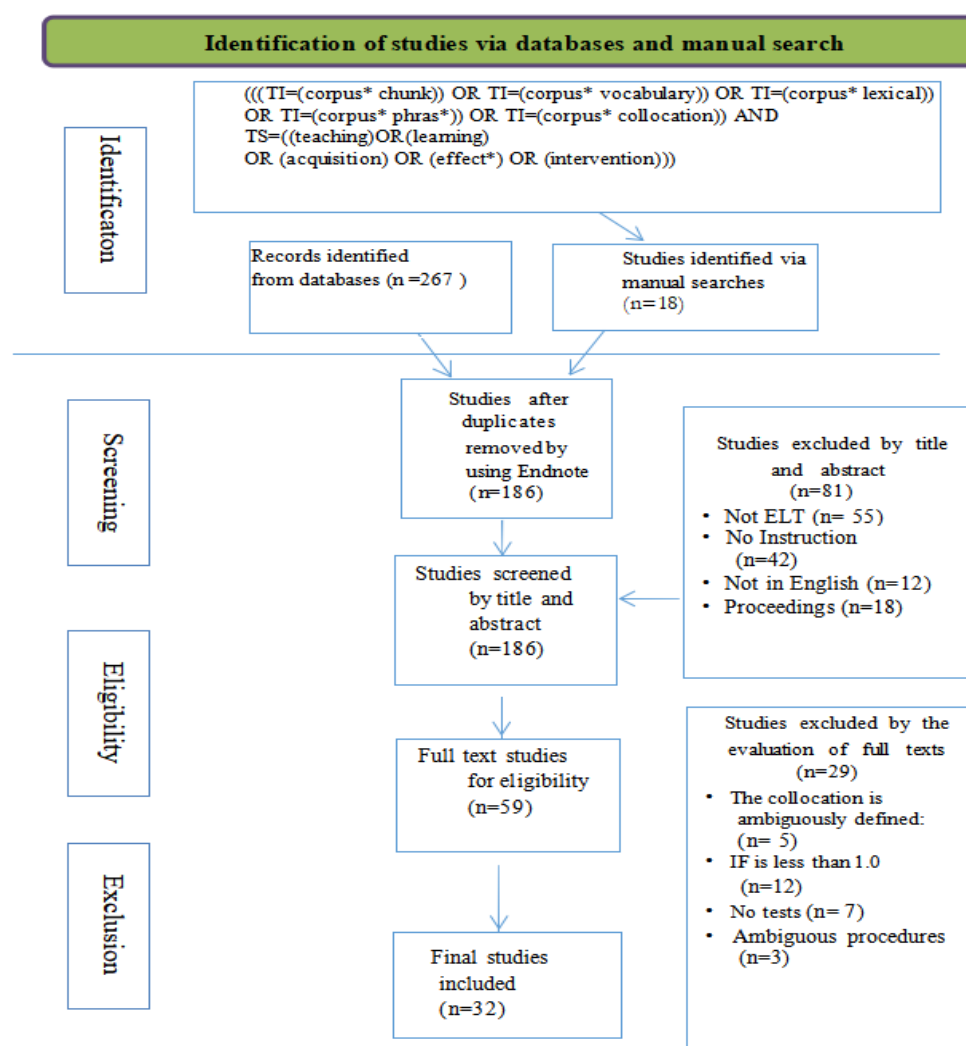
Identification of studies via databases and manual search

(((TI=(corpus* chunk)) OR TI=(corpus* vocabulary)) OR TI=(corpus* lexical)) OR TI=(corpus* phras*)) OR TI=(corpus* collocation)) AND TS=((teaching)OR(learning) OR (acquisition) OR (effect*) OR (intervention)))

Records identified from databases (n =267)

Studies identified via manual searches (n=18)

Studies after duplicates removed by using Endnote (n=186)

Studies excluded by title and abstract (n=81)
- Not ELT (n= 55)
- No Instruction (n=42)
- Not in English (n=12)
- Proceedings (n=18)

Studies screened by title and abstract (n=186)

Full text studies for eligibility (n=59)

Studies excluded by the evaluation of full texts (n=29)
- The collocation is ambiguously defined: (n= 5)
- IF is less than 1.0 (n=12)
- No tests (n= 7)
- Ambiguous procedures (n=3)

Final studies included (n=32)

Identification | Screening | Eligibility | Exclusion

Figure 1. The Search Process Flow Diagram Showing the Use of PRISMA 2020.

Due to the fact that some researchers mixed collocation with other terms, such as lexical expression, and chunk and phrase, some important papers would have been excluded if we confined the search word to "collocation". Thus, the searching keywords used in the literature search were corpus*, chunk/vocabulary/lexical/phras*/concordanc*, instruct*/teach*/intervention/efficacy, and second/foreign language/Data-driven learning, and collocation.

3.3. Phase 2: The Screening Phase

When conducting searches in the databases and Google Scholar, duplicate entries were identified using EndNote 20 and subsequently removed. Subsequently, the article titles were re-evaluated to verify that the remaining entries aligned with the criteria established by the authors of this review.

3.4. Phase 3: The Eligibility Phase

During the third phase, the collected articles underwent an eligibility assessment, necessitating their alignment with the inclusion criteria outlined in the inclusion section in Table 2. It is noteworthy that this phase carried significance in ensuring a study's data were of high quality and reliability.

Table 2. Inclusion and Exclusion Criteria.

Criterion	Inclusion	Exclusion
Type of article	Journal articles and education-related	Book, book chapter, systematic review, proceedings
Language	English	Non-English
Year	2003–2022	<2003
Peer review	Peer-reviewed	Non-peer-reviewed
Methodology	Quantitative, qualitative, mixed method	Text analysis, ambiguously described
Term defined	Consistent with selected	Inconsistent
Instruction	Experimental or case study	Not specified

3.5. Phase 4: Exclusion

We searched several databases and journals for studies published in the last two decades (2003–2022) due to the recent fast development trend in computer-aided education. We excluded non-research-based articles, such as blogs, proceedings, editorials, and book reviews. Studies published in non-English journals were also removed during the search stage.

In addition, other publications were also excluded from this systematic literature review after the third phase of eligibility checks. After reading the basic information in the papers, we excluded articles that were published before 2003. Along with poorly conceived research, we also made sure that non-peer-reviewed articles were eliminated, and removed those with ambiguously described terms. Studies with no specified instructions/interventions were deleted. Low-quality papers in some journals (IF < 1.0) were also excluded. The exclusion phase was crucial to ensuring that we obtained high-quality data, similar to the eligibility phase. Figure 1 provides a more detailed entire procedure.

The selected papers were published in SSCI or Scopus journals on English Education, such as *Applied Linguistics*, *Applied Linguistics Review*, *Asian EFL Journal*, *CALICO Journal*, *English for Specific Purposes*, *Eurasian Journal of Applied Linguistics*, *International Journal of Applied Linguistics (United Kingdom)*, *International Journal of Applied Linguistics and English Literature*, *Journal of Asia TEFL*, *Journal of English for Academic Purposes*, *Language Learning and Technology*, *System*, *TESL-EJ*, *TESOL Journal*, *TESOL Quarterly*. If the paper was not in the above-mentioned journals, it was onlyincluded if it was published in a high-quality journal (IF > 1.0).

3.6. Data Analysis and Coding

The content analysis method [48] was applied during the data analysis part of this study. This approach is frequently used to evaluate text, make comparisons, and groupdata into categories. The first step involved creating a table for recording the findings of the analysis performed using MS Word. Categories that are connected to our research questions make up the structure, for instance, the study's corpus tool and the environment in which it was conducted. Each article was thoroughly read, and information pertinent to filling out the table was gathered. Following the completion of the questionnaire, the next phase involved utilizing MS Excel 365 to assign codes, as shown in Table 3, and categories as a means of analysis.

Table 3. Data Categories and Codes.

RQ	Dimension	Subdimension	Code
RQ 1	Context	Ages of Participants	**age ≥18 coded into "adult".** **age between 14 and17 coded into "teenager".** **Age-not-mentioned college students coded into "adult".** Age-not-mentioned secondary school students coded into "teenager".
		Sample Size	Numeric data
		Level/proficiency	Preliminary Low-intermediate Pre-intermediate Intermediate, B1 = intermediate Upper-intermediate, B2 = upper intermediate advanced TOEIC 329-990 = low-intermediate to advanced Not specified = NP
		Territory	Country names
		Institute	Colleges and universities coded as "tertiary school". Low- and high-middle schools coded as "secondary" school. adult participant with no institute coded as "tertiary school".
RQ 2	Corpora and tools	Web-based	COCA, SKELL, BNC, LEXTUTOR, etc.
		Offline software/App	Antconc, Digital-dictionary, Wordsmith 6, etc.
RQ 3	Research design	Instructional method	Direct, indirect, implicit, explicit, inductive, deductive, hard, soft, symmetric, asymmetric, scaffolding.

4. Findings

Following completion of the four selection phases for eligible articles, a total of 32 full-text, peer-reviewed articles were deemed suitable for inclusion in this review. As a result, this section will comprehensively delve into a detailed discussion of the outcomes from the previously selected studies.

4.1. Research Contexts Regarding the Corpus-Based Learning Approach in EFL Learners' English Collocation Acquisition

As shown in Table 4, the first research question was addressed using the distribution of the studies by year, geography, characteristics of the individuals recruited into the chosen publications, and samples used in the reviewed studies.

Table 4. Research Context Information.

No.	Author and Year	Country	Institution	Level/Proficiency	Ages	Sample Size
1	[49]	Iran	tertiary school	intermediate	adults	75
2	[50]	Turkey	tertiary school	advanced	adults	58
3	[41]	Malaysia	tertiary school	not specified	adults	60
4	[51]	Saudi Arabia	tertiary school	beginner-to-intermediate	adults	62
5	[52]	Saudi Arabia	tertiary school	intermediate	adults	51
6	[53]	Iran	tertiary school	upper-intermediate to advanced level	adults	44
7	[54]	Iran	tertiary school	intermediate	adults	60
8	[39]	China	tertiary school	not specified	adults	32

Table 4. *Cont.*

No.	Author and Year	Country	Institution	Level/Proficiency	Ages	Sample Size
9	[55]	Republic of Macedonia	tertiary school	not specified	adults	54
10	[56]	Turkey	secondary school	not specified	teenagers	42
11	[36]	Turkey	tertiary school	upper-intermediate	adults	70
12	[57]	Japan	tertiary school	not specified	adults	26
13	[58]	China	tertiary school	upper-intermediate	adults	40
14	[59]	Iran	tertiary school	not specified	adults	84
15	[60]	Malaysia	tertiary school	not specified	adults	40
16	[61]	Korea	tertiary school	intermediate to advanced	adults	48
17	[62]	China	tertiary school	not specified	adults	60
18	[63]	Poland	secondary school	not specified	teenagers	60
19	[15]	Iran	tertiary school	various levels	adults	200
20	[64]	China	tertiary school	not specified	adults	52
21	[65]	Iran	tertiary school	upper-intermediate	adults	45
22	[66]	Iran	tertiary school	intermediate	adults	125
23	[67]	Iran	tertiary school	not specified	adults	60
24	[68]	Iran	tertiary school	intermediate	adults	100
25	[43]	Iran	secondary school	low-intermediate	teenagers	60
26	[69]	Iran	tertiary school	intermediate	adults	80
27	[70]	Japan	tertiary school	intermediate	adults	55
28	[40]	China	secondary school	not specified	teenagers	81
29	[71]	Turkey	tertiary school	pre-intermediate	adults	30
30	[72]	China	tertiary school	intermediate to upper-intermediate	adults	72
31	[73]	Korea	tertiary school	low-intermediate to upper-intermediate	adults	115
32	[74]	Turkey	tertiary school	pre-intermediate	adults	30

4.1.1. Distribution of Articles Based on Time

As shown in Figure 2, the overall, trend suggests progressive growth in the research and publications related to corpus-based instruction of collocation over the past two decades, reflecting the increasing recognition of its significance and potential in language education. In 2005, corpus-based instruction of collocation started gaining attention, with publications addressing its effectiveness and pedagogical implications. From 2008 to 2014, there was a gradual increase in publications related to corpus-based instruction of collocation, indicating a growing interest in the field. The years 2015 and 2016 witnessed a significant surge in publications, suggesting a heightened focus on corpus-based instruction of collocation during this period. In 2017, there was a notable presence of publications, indicating a sustained interest and ongoing research in the field. From 2019 to 2022, there continued to be a steady flow of publications, indicating a sustained interest and relevance of corpus-based instruction of collocation in language teaching and learning. Overall, the research on corpus-based instruction of collocation in language teaching and learning has seen a steady rise from 2005 to 2022.

This description implies that there has been consistent and substantial growth in research and publications concerning corpus-based instruction of collocations over the last twenty years. The increasing trend reflects a growing awareness of the importance and potential of this approach in language education.
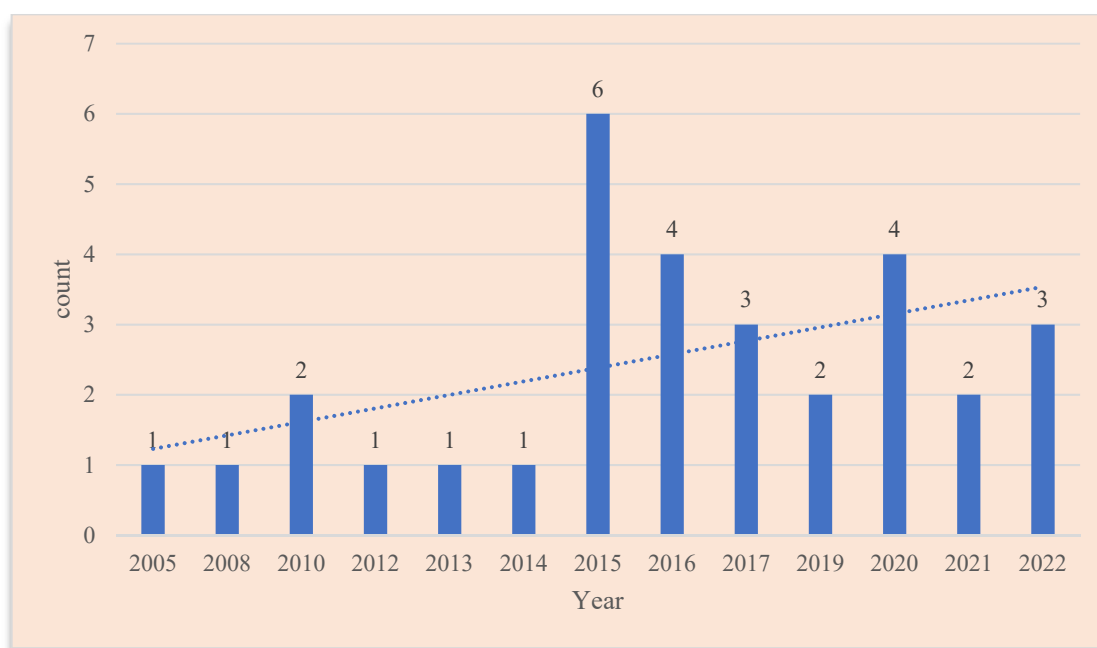
Figure 2. Distribution of Articles Based on the Timeline Considered.

4.1.2. Distribution of Articles Based on Regions

The distribution of articles based on territory, as shown in Figure 3, indicates that there is a diverse range of regions and countries where corpus-based collocation instruction was studied or implemented. Iran has the highest count with 11 instances (e.g., [13,43–45]), followed by China [39,40,58,62,64,72] and Turkey [36,50,56,71,74], with 6 and 5 instances, respectively. Other countries such as Japan [57,70], Korea [61,73], Saudi Arabia [51,52], Yemeni [41], Malaysia [60], Poland [63], and the Republic of Macedonia [55] have fewer instances, ranging from two to one.
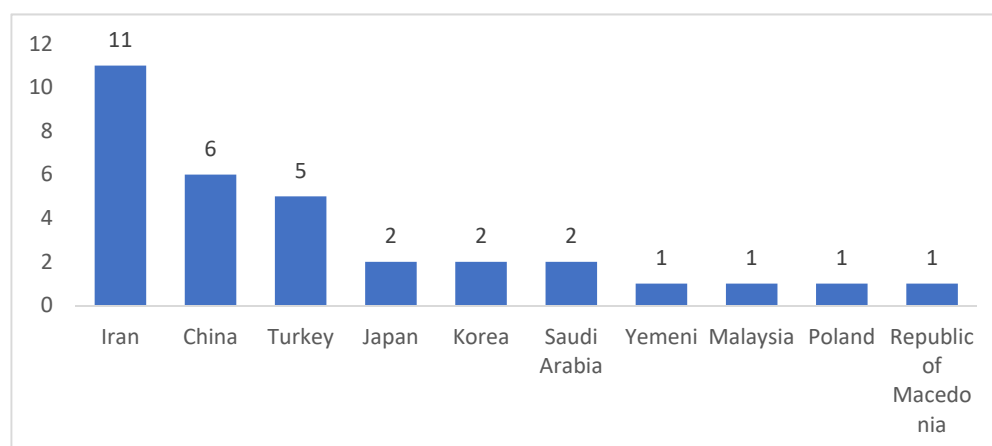
Figure 3. Distribution of Articles Based on Territory.

This distribution suggests that corpus-based collocation instruction has gained attention and has been researched or implemented mainly in Asian regions.

4.1.3. Distribution of Research Institutions

Based on the information shown in Figure 4, research institutions can be categorized into two types: secondary schools (4 studies) and tertiary schools (28 studies). This indicates that the majority of corpus-based collocation instruction research has been conducted in tertiary school settings, with 28 studies (e.g., [39,41,50–53,55,57]) dedicated to this level of

education. However, there are also four studies (e.g., [40,43,56,63]) that focus on secondary schools, demonstrating some attention to collocation instruction at that level as well.
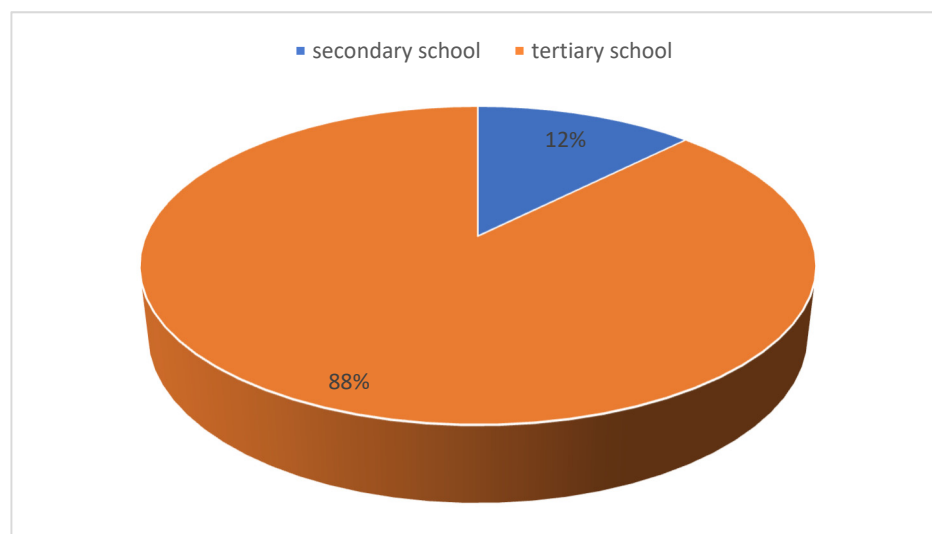

Figure 4. Distribution of Research Institutions.

4.1.4. Participants' Language Proficiency

The participants' language proficiency in each study was described based on various criteria and was matched to the CEFR level by the authors of this review. It is important to note that the proficiency level descriptions in Figure 5 are diverse and may vary depending on the specific framework or context. Some instances include specific levels like advanced [50], pre-intermediate [71,74], upper-intermediate [36,58,65,68], and intermediate [26,52,54,66,69,70], while others have more general descriptions such as beginner-to-intermediate [51], intermediate-to-upper-intermediate [72] However, some studies provided no information except the participants' college grades [39–41,55–57,59,60,62–64,67]. This distribution indicates that corpus-based collocation instruction was investigated across a range of proficiency levels, catering to learners at different stages of language proficiency.
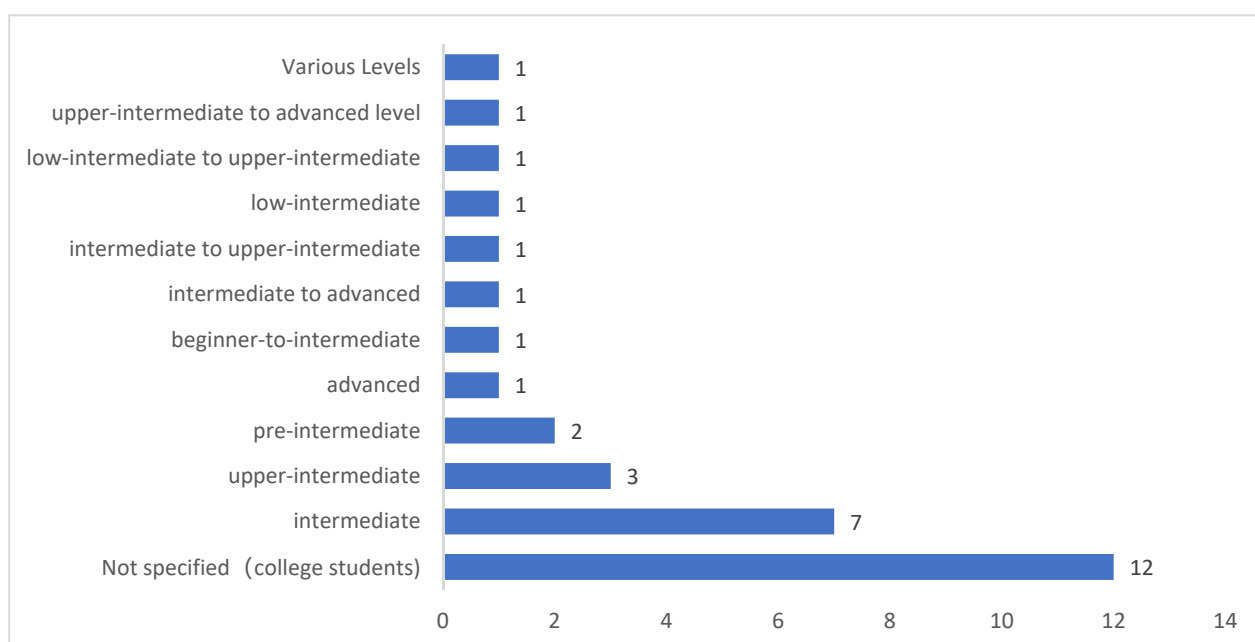

Figure 5. Participants' Language Proficiency.

4.2. Type of Corpus Utilized to Improve ELF Learner's Collocation

Table 5 lists a variety of tools used for collocation teaching in the reviewed studies. The web-based mode is the most prevalent, with the BNC and COCA corpora being frequently utilized. Web-based tools and resources allow for easy access to large language corpora and concordance tools. Offline modes involve the use of software, such as AntConc, or printed collocation dictionaries. Mixed modes (e.g., [61,73]) involve combinations of different corpus tools. What is more, one study involved a self-built corpus [58]. However, one study [54] failed to provide specific information about the corpus used in the intervention.

Table 5. Distribution of Corpus Tools.

Mode	Count	Corpus/Tool	Author and Year
Web-based	6	BNC	[50,55,56,59,63,70]
	8	COCA	[36,41,51,53,62,64,71,72]
	1	TerminoWeb	[49]
	1	Law of Contract Corpus, BNC	[60]
	5	Concordancer	[15,39,40,67,68]
	1	SKELL	[57]
Offline	3	AntConc	[43,52,74]
	3	Collocation Dictionary	[65,66,69]
	1	Teacher-built	[58]
Mixed	1	BNC, VLC Web Concordancer, Google	[61]
	1	BNC, COCA, LEXTUTOR, WORDSMITH6.0	[73]
Not Specified	1	A Corpus	[54]

It is worth noting that the mentioned tools may represent a subset of the overall landscape of collocation teaching tools and modes, as they are specific to the reviewed studies.

4.3. Types of Instructional Methods Used

The hands-on and hands-off approaches used in the experimental groups in these studies are displayed in Figure 6. There are seven studies [15,49,54,58,59,69,71,73] in which a hands-off approach is used, accounting for 22% of the total. The hands-off approach involves learners observing and analyzing collocations without direct interaction. It may include activities such as reading and analyzing corpus data or examining collocation examples. There are 21 instances (66%) (e.g., [36,39,41,51–53,55–57]) in which a hands-on approach is used. The hands-on approach emphasizes active engagement and learner participation. It involves interactive tasks and exercises that require learners to actively use and manipulate collocations. Examples include completing collocation exercises, producing collocations in context, and engaging in collaborative activities. There are two studies [40,43] in which both hands-on and hands-off approaches are used and compared, suggesting various activities that involve both active engagement and passive observation. There are two instances [43,60] where a combination of hands-on and hands-off approaches is explicitly mentioned. These findings demonstrate the different approaches used in teaching collocations based on corpora. While some studies emphasize hands-on activities in which learners actively manipulate and use collocations, others use a more hands-off approach that focuses on observation and analysis.
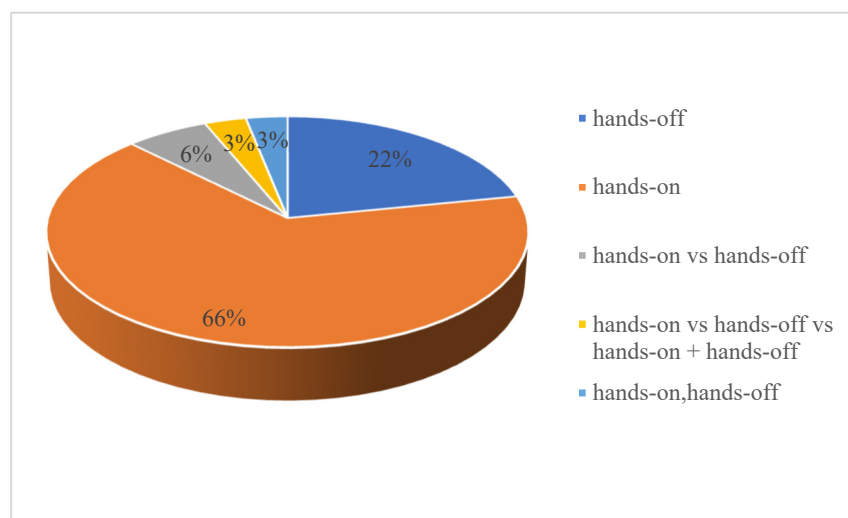
Figure 6. Distribution of Instructional Methods.

4.4. Impact of Corpus-Based Instruction on EFL Learners' Collocation Learning

The impact of corpus-based instruction on collocation in each study, found in both the experimental results and survey findings are listed in Table 6. These studies explored the effectiveness of corpus-based tools and data-driven learning in teaching collocations to EFL learners of various levels, among which 27 studies report significant effects of corpus-based instruction with different methods (direct, indirect, inductive, deductive, output-based, and input-based). For instance, Akbari et al. [49] discovered that corpus-based tools significantly impacted how university students were taught collocations in various specialized disciplines. Al-Mahbashi et al. [41] and Almegren [51] demonstrated that DDL groups outperformed dictionary groups in terms of learning outcomes. Ashouri et al. [54], Li [62], and Łuszcz [63] highlighted the benefits of corpus-based instruction in improving learners' collocation knowledge and usage. Huang [58] found that concordance activities improved learners' productive language by enhancing accuracy and complexity. Jafarpour et al. [59] and Koosha and Jafarpour [15] reported the effectiveness of concordance-based methods and DDL in teaching collocations. Men [64] emphasized the positive impact of DDL on collocation production, while Mohammadi and Khalaji [65] found that learners had positive perceptions of corpus-based design. Other studies, such as Yunus and Awab [60], Rezaee et al. [68], Saeedakhtar et al. [43], and Yang [73] highlighted the advantages of DDL and concordance activities using questionnaire responses and improved performance on collocation tasks. Additionally, Sun and Wang [40] found that inductive approaches were more effective for easy collocations, while Uçar and Yükselir [71] demonstrated the significant impact of corpus-based activities on learning verb-noun collocations. Wu [72] reported sustained improvement in learners' performance on change-of-state verbs following DDL instruction. Lastly, Yilmaz [72] indicated that an experimental group exhibited a greater variety of collocational patterns and fewer linguistic errors, with positive attitudes toward DDL and concordance activities. In summary, these studies provide evidence supporting the effectiveness of corpus-based instruction and DDL in teaching collocations, leading to improved collocation knowledge and usage among language learners.

In addition, several studies investigated the long-term effects of corpus-based collocation instruction. From the perspective of the long-term effect of corpus-based collocation instruction (DDL), several studies provide valuable insights. Akbari et al. [49] demonstrated that corpus-based tools significantly influenced the teaching of collocations, with lasting effects observed over three weeks. Al-Mahbashi et al. [41] found that a DDL group outperformed a dictionary group, and this superiority was maintained even in a delayed post-test. Chan and Liou [39] showed that web-based concordance instruction initially improved participants' collocation skills, although regression occurred later. Durmus and Kikimci [56] reported that deductive DDL instruction was more effective than inductive

instruction for learning and retaining target collocation patterns. Wu [72] reported that DDL instruction led to long-lasting improvements in the performance of change-of-state verbs. Lastly, Yang [73] showed that collaborative corpus-based learning benefited students of all proficiency levels, leading to the long-term retention of collocation items and their usage. Overall, these findings indicate the enduring benefits of corpus-based collocation instruction (DDL) on learners' retention, awareness, and usage of collocations.

What is more, two studies incorporated different methods and tasks in the design of their methods. Rezaee et al. [68] observed significant improvements in collocation knowledge for participants in asymmetrical and symmetrical concordance tasks with scaffolding conditions. The learners exhibited favorable attitudes regarding scaffolding and concordancing. Yang [73] found that collaborative corpus-based learning (CCL) was more effective for high-level students in long-term collocation retention, with no significant difference between CCL and individual corpus-based learning (ICL) groups in collocation awareness.

However, the findings of some studies present conflicting results regarding the effectiveness of implicit and explicit instruction methods. According to the findings of Saeedakhtar et al. [43], Durmus and Kilimci [56], and Sun and Wang [40], explicit instruction was shown to produce more favorable results. However, Akinci and Yildiz [50] discovered that DDL (implicit) utilizing corpus consultation was not as effective as explicit (deductive) instruction. In a similar vein, Yunus and Awab, [60] recommended the explicit teaching of prepositional collocations using both "soft" and "hard" DDL approaches rather than relying solely on independent or "hard" DDL instruction.

Here we describe the additional findings of corpus-based collocation instruction from Altun [53] and Salehi & Rasooyar [69]. Altun found that the corpus-based approach had a greater impact on the reception (understanding) of strong collocations. This suggests that learners were able to comprehend and recognize strong collocations more effectively with corpus-based instruction. On the other hand, the acquisition (learning) of weak collocations was better for the group that used dictionaries as a resource. This indicates that learners benefited more from using dictionaries to learn and remember weaker collocations. Surprisingly, Salehi and Rasooyar's [69] study showed that corpus-based training had the same impact on EFL learners' acquisition of incongruent collocations as traditional instruction. This means that both methods, corpus-based instruction and traditional instruction, were equally effective in teaching non-congruent collocations to EFL learners.

These findings provide valuable insights into the effectiveness of corpus-based instruction in different aspects of collocation learning, such as the reception of strong collocations and the acquisition of weak collocations. They also highlight the comparable effectiveness of corpus-based instruction with traditional instruction in teaching non-congruent collocations.

Table 6. Summary of Findings.

No.	Author	Instruction Method	Results and Findings
1	[49]	hands-off	In the short and long term (three weeks), corpus-based tools had a considerable impact on teaching collocations to university students with various specialized fields.
2	[50]	hands-on vs. hands-off vs. hands-on + hands-off	The study revealed that DDL, which involved consulting corpora, was less effective compared with explicit instruction. Nevertheless, the results of a self-evaluation questionnaire highlighted a positive view among students toward DDL. The questionnaire also indicated that the incorporation of corpora was perceived as more advantageous and influential.
3	[41]	hands-on	The outcome of the delayed post-test demonstrated a noteworthy improvement in learning for the DDL group in comparison with the dictionary group. The DDL group exhibited significantly superior learning outcomes compared with the dictionary group.
4	[51]	hands-on	The DDL group outperformed the dictionary group.

No.	Author	Instruction Method	Results and Findings
5	[52]	hands-on	Tracking logs and activity sheets showed that participants were able to use DDL and benefited from it.
6	[53]	hands-on	Strong collocations were more easily absorbed using a corpus-based technique, while weak collocations were more easily learned by those who used dictionaries.
7	[54]	hands-off	Compared with traditional training, corpus-based collocation instruction helped learners learn and use collocations more effectively.
8	[39]	hands-on	Participants' collocation was improved with web-based concordance teaching immediately following online exercises, but this improvement eventually declined. According to a survey, the concordancer and online educational units were well-liked by the majority of participants.
9	[55]	hands-on	Instruction using web-based concordance led to increased knowledge and better performance across all test sections compared with traditional activities.
10	[56]	hands-on	The target collocation patterns were much easier to learn and remember for the DDL group that received deductive (explicit) teaching as opposed to inductive (implicit instruction).
11	[36]	hands-on	Corpus-based activities assisted students in identifying, comprehending, and generating accurate forms of phrasal–prepositional verbs. Nevertheless, these activities were not successful in facilitating students' comprehension of the metaphorical meanings associated with these structures.
12	[57]	hands-on	The results obtained from a questionnaire suggested a strong correlation between how students perceive the utilization and effectiveness of the resource and their attitudes toward English education.
13	[58]	hands-off	The EG outperformed the CG with a broader range of collocational patterns and fewer grammatical errors when using specified abstract nouns. Survey results highlighted that participating in concordance exercises enhanced a practical learning approach, helping students recognize lexical and prepositional collocations. This led to improved accuracy and complexity in their productive language skills.
14	[59]	hands-off	The use of a concordance-based method had a greater impact on the understanding of collocations among L2 learners compared with the traditional approach.
15	[60]	hands-on, hands-off	The DDL group outperformed the comparison group notably in tasks involving sentence completion and semantic function identification. The study's authors recommended instructing prepositional collocations explicitly using both "soft" and "hard" DDL methods rather than relying solely on independent or "hard" DDL approaches.
16	[61]	hands-on	The implementation of a corpus DDL approach had a substantial impact on increasing college students' awareness of vocabulary and grammar, particularly in relation to various types of collocations. The results of a survey indicated that the corpus DDL approach promoted the development of language learning autonomy among college students.
17	[62]	hands-on	The students assigned to the experimental group exhibited a considerable advancement in utilizing collocations. The acquisition and utilization of corpora by students contributed to enhancing their understanding of customary collocational usage and developing their competence in using collocations.
18	[63]	hands-on	The findings suggested that students who had exposure to corpora demonstrated improved ability in handling challenges related to the use of collocations in English, particularly in speaking activities.
19	[15]	hands-on	The DDL approach demonstrated a high level of effectiveness in instructing and acquiring collocations of prepositions compared with the conventional instruction group. The learners' performance in mastering prepositional collocations was found to have a positive correlation with their proficiency level.

No.	Author	Instruction Method	Results and Findings
20	[64]	hands-on	Under the DDL model, there was a notable enhancement in learners' ability to produce collocations, whereas the use of dictionaries did not demonstrate a similar influential impact. The participants' evaluations in a questionnaire regarding the usefulness of dictionaries were not as favorable as those concerning the online corpus. It was concluded that DDL offers valuable advantages for exploratory learning.
21	[65]	hands-on	The findings revealed that the experimental groups had a notable advantage in acquiring collocations and applying them in their written work. Positive opinions of the corpus-based design were evident in interviews with seven students, despite the mentioned limitations.
22	[66]	hands-on	The learners used inputoutput and corpus-based instructions as more effective strategies to address the challenge of comprehending collocations.
23	[67]	hands-on	The EG outperformed the CG on the post-test, indicating that collocation training could improve students' language ability.
24	[68]	hands-on	Participants exposed to concordancing tasks under different conditions (symmetrical, asymmetrical, and no-scaffolding) showed substantial enhancement in understanding collocations for both receptive and productive tasks in contrast with the control group. Yet, no significant distinctions emerged among the concordancing groups in task performance.
25	[43]	hands-on vs. hands-off	The results demonstrated enhanced performance of the experimental groups in acquiring verb–preposition collocations on the immediate post-test compared with the control group. There was no noteworthy distinction between the experimental groups. Nevertheless, the hands-on group exhibited superior retention of knowledge on the delayed post-test compared with the hands-off group. The questionnaire outcomes emphasized the favorable view of learners toward DDL for acquiring collocations.
26	[69]	hands-off	The outcomes revealed that corpus-based and traditional instruction had comparable impacts on EFL learners' acquisition of non-congruent collocations. Yet, students displayed favorable views of corpus-based instruction and favored it over the conventional teaching method for collocations.
27	[70]	hands-on	Corpora utilization yielded more substantial collocation outputs compared with dictionary usage.
28	[40]	hands-on vs. hands-off	The results showed that the inductive (implicit) group excelled over the deductive (explicit) group in collocation learning, particularly with simpler collocations with concordancers. Furthermore, the study revealed that the complexity of grammatical patterns did not notably impact the performance disparity between the inductive and deductive methods. This suggested equal effectiveness for challenging items in both approaches.
29	[71]	hands-off	The study found that instruction using corpus-based activities (concordancing) had a significant impact on students' learning of verbnoun (V-N) collocations compared to the dictionary group.
30	[72]	hands-on	Using DDL led to enhanced performance in change-of-state verbs, and this enhancement was maintained even after a three-month treatment period.
31	[73]	hands-off	CCL (collaborative corpus-based learning) excelled over ICL (individual corpus-based learning) among high-level students. Irrespective of language proficiency, CCL enhanced long-term retention and usage of collocations. Both CCL and ICL groups similarly improved collocation awareness. High-level students showed more favorable attitudes and perceptions than low-level students.
32	[74]	hands-on	The findings revealed that the experimental group used a broader variety of collocational and colligational patterns than the control group and made fewer grammatical mistakes while utilizing abstract nouns. The survey results showed that students' attitudes toward the usage of DDL and concordance activities were overwhelmingly positive. Additionally, they stated a desire to participate in DDL activities in the future.

5. Discussion

The first question asked in this review was about the research contexts related to the corpus-based learning approach in the acquisition of English collocations by EFL learners. During this period, many studies and publications emerged that explored the effectiveness of corpus-based vocabulary instruction. Over the past two decades, academics and educators have paid a great deal of attention to data-driven learning, which involves the direct or indirect application of corpus technology [75]. It can be argued that advancements in technology have made it more accessible for educators to incorporate corpus-based resources and tools into their teaching practices [76–80]. Hanafiyeh [80] thought that as technology advanced, corpus-based language acquisition became the emphasis of L2 teachers and scholars. Online corpora, concordancers, and specialized software have become more user-friendly and readily available, enabling language teachers to utilize corpus-based instruction in their classrooms [68]. Click or tap here to enter text.

The distribution of research institutes revealed that corpus-based collocation instruction was primarily conducted in tertiary school settings while also receiving attention in secondary schools. This suggests investigations across proficiency levels, offering insights into the effectiveness of corpus-based collocation instruction for learners at different stages of language development. This backs up the claim made in the literature that collocations are beneficial for improving students' accuracy and fluency as well as their overall language proficiency at all levels [8,81,82]. This finding aligns with a study conducted by Youmei and Yun [83], which highlighted the challenges that even advanced EFL learners face when it comes to producing collocations. The use of collocations in second language learning is commonly recognized to have a considerable impact, especially at the intermediate and advanced levels compared with the usage of collocations in the native language [54]. In addition, the presence of four studies focusing on corpus-based collocation instruction in secondary schools suggests that this approach is feasible even among students with lower language proficiency levels. These studies provide evidence of the potential effectiveness of corpus-based vocabulary teaching in supporting the language development of secondary school students at various proficiency levels [84].

The various corpora and tools used in these studies highlight the range of tools available for collocation teaching. The preference for online resources is attributed to their accessibility and the availability of vast language corpora and concordance tools. According to Boulton [85] and Chambers [86], the availability and user-friendliness of concordancers have made it easier for learners to access and utilize large English reference corpora. Offline modes, including software like AntConc and printed collocation dictionaries, were also utilized. What is more, some studies [56,58] used mixed modes, combining different corpus tools while others explored self-built teaching tools [53].

Hands-on approaches significantly outnumbered hands-off approaches in the reviewed studies. These approaches highlight the active role of teachers in harnessing corpus technology to enhance language learning and instruction. The diversity of collocation concordance tools discussed in the findings led to the identification of two main teaching methods: hands-on and hands-off. As shown in Table 6, over the past two decades, research on students' direct use of corpora in language learning has increased since Johns [38] advocated DDL in language teaching and learning [58]. Furthermore, the use of online tools and resources, such as the BNC and COCA corpora, enables a hands-on approach, allowing students to actively engage with the data and interact with the concordance tools [85,86]. On the other hand, offline modes, such as software or printed dictionaries, use a hands-off approach, where students rely on pre-compiled information without direct interaction with the language corpus. These two teaching methods offer different levels of student engagement and autonomy in the exploration and understanding of collocations.

The effectiveness of corpus-based instruction and data-driven learning in teaching collocations to EFL learners was examined in the reviewed studies. The results showed that corpus-based instruction had a significant positive effect on collocation learning, with different methods and approaches yielding successful outcomes across students of diverse

levels. These results are consistent with those of earlier studies that sought to determine whether using corpora would help language learners learn new vocabulary (e.g., [87–89]), more specifically, collocation learning.

DDL was effective in enhancing learners' collocational competence, leading to better retention and performance in the long term. Some pedagogical theories underpin the effectiveness of corpus-based hands-off and hands-on approaches in collocation instruction, enhancing learners' understanding and acquisition of collocations [66]. The first one is rooted in the noticing hypothesis [90] and grammatical consciousness raising [91]. The second function is associated with learning processes such as discovery learning, and inductive learning as well as generalization. These theories fundamentally direct learners' focus to language forms, whether using authentic language instances or pedagogically generated data.

Collaborative corpus-based learning and mixed procedures of explicit instruction and output are also beneficial. The findings in the reviewed papers corroborate those of Minaei and Rezaie [92], who found that collaborative output tasks resulted in more collocation knowledge than individual output tasks. Communicative activities with corpus-based collaborative learning should be considered and arranged for classes [85,86], because collaborative learning stems from communicative language teaching (CLT) as a teaching and learning approach [93], and group collaboration may provide students with greater opportunities to be active participants in creating and applying usages [73].

However, such studies as Saeedakhtar et al. [43], Durmus and Kilimci [56], and Sun and Wang [40] presented conflicting findings regarding the effectiveness of implicit and explicit instruction methods, suggesting that different instructional approaches may yield varying results. This is a continuation of contention between explicit and implicit instructional approaches. Additionally, the study [53] highlighted the reception and acquisition of strong and weak collocations, indicating the impact of corpus-based instruction and the complementary role of dictionaries in learning collocations. Furthermore, the comparable effectiveness of corpus-based instruction and traditional instruction in teaching non-congruent collocations was observed [69]. These findings provide valuable insights into the positive impact of corpus-based instruction on collocation learning and highlight the effectiveness of different instructional methods and approaches in enhancing learners' collocational competence.

6. Conclusions

This analysis of the distribution of articles based on various factors provides valuable insights into the trends and characteristics of corpus-based collocation instruction in language teaching and learning. The findings demonstrate a growing trend in this topic over the past two decades, indicating an increasing recognition of its significance and potential in language education. The distribution based on time highlights the heightened focus on corpus-based collocation instruction in recent years, with sustained interest and ongoing research in the field.

Furthermore, corpus-based collocation instruction was extensively researched and implemented in Asian regions, as evident from the distribution based on regions. Research in this area is primarily conducted in tertiary school settings, but attention is also given to collocation instruction at the secondary school level. Different proficiency levels are targeted in corpus-based collocation instruction research, catering to learners at various language development stages. The use of diverse corpus tools, both online and offline, highlights the flexibility and accessibility of resources.

Research on corpus-based instruction and its impact on collocation learning consistently demonstrates its effectiveness. Various methods were shown to improve students' collocation knowledge, usage, and long-term retention. However, conflicting results regarding implicit and explicit instruction methods call for further investigation. Factors influencing instructional effectiveness in specific contexts need to be explored for more

nuanced guidance. Traditional instruction and corpus-based instruction were found to be similarly effective for teaching non-congruent collocations.

Overall, the reviewed studies collectively demonstrate the positive impact of corpus-based instruction on learners' collocation knowledge, usage, and retention. The findings contribute valuable insights into the effectiveness of different instructional methods and highlight the need for further research to optimize collocation instruction strategies.

This research synthesis offers both theoretical and pedagogical implications. First, by systematically searching, selecting, and analyzing relevant studies, we synthesized the existing evidence in a structured and comprehensive manner to provide a clear overview of the current state of knowledge on the topic. Second, by analyzing and comparing the findings of multiple studies, we assessed the impact of this instructional approach on learners' collocational competence. This evaluation helps identify the strengths, weaknesses, and limitations of corpus-based collocation instruction and provides insights into its effectiveness. Third, this study may help identify gaps in the existing research on corpus-based collocation instruction. By critically assessing the literature, we determined areas where further research is needed. This identification of research gaps guides future research efforts and helps prioritize areas that require additional investigation or exploration. The findings of this systematic review may provide evidence that can inform educational policies, curriculum development, and instructional practices.

Academics interested in this topic can address this study's limitations in the future. Although we searched multiple databases, there might be other relevant databases or sources of information that were not included in the search. This could lead to a potential bias in the selection of studies and may result in the exclusion of relevant articles. The inclusion criterion specifies that only English-language articles were considered. This could lead to language bias, as relevant studies published in other languages may have been excluded. Language bias can limit the diversity of perspectives and findings included in a synthesis, potentially affecting the generalizability of the results. Finally, it is necessary to examine the effect sizes of these empirical studies and conduct a meta-analysis, which will enhance rigor, validity, and generalizability, providing a more comprehensive understanding of the treatment effect or relationship under investigation.

In summary, the systematic review of corpus-based collocation instruction is essential for synthesizing the evidence, evaluating effectiveness, identifying best practices, and informing policy and practice. This review serves as a valuable tool for evidence-based decision-making, knowledge generation, and improvement of language teaching approaches.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data created by the systematic research are present in the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Benson, M. Collocations and Idioms. In *Dictionaries, Lexicography and Language Learning*; Pergamon Press: Oxford, UK, 1985; Volume 1, pp. 61–68.
2. Cowie, A.P.; Cater, R.; McCarthy, M. Stable and Creative Aspects of Vocabulary. In *Vocabulary and Language Teaching*; Longman: New York, NY, USA, 1988; Volume 139, pp. 126–129.

3. Lewis, M.; Gough, C.; Martínez, R.; Powell, M.; Marks, J.; Woolard, G.C.; Ribisch, K.H. *Implementing the Lexical Approach: Putting Theory into Practice*; Language Teaching Publications: Hove, UK, 1997; pp. 223–232.

4. Hashemi, M.; Azizinezhad, M.; Dravishi, S. Collocation a neglected aspect in teaching and learning EFL. *Procedia-Soc. Behav. Sci.* **2012**, *31*, 522–525. [CrossRef]

5. Ellis, N.C. Sequencing in SLA: Phonological memory, chunking, and points of order. *Stud. Second Lang. Acquis.* **1996**, *18*, 91–126. [CrossRef]

6. Leffa, V.J. Textual constraints in L2 lexical disambiguation. *System* **1998**, *26*, 183–194. [CrossRef]

7. Ellis, N.C. Memory for language. In *Cognition and Second Language Instruction*; Robinson, P., Ed.; Cambridge University Press: Cambridge, UK, 2001; pp. 33–68.

8. Nation, I.S.P. *Learning Vocabulary in Another Language*; Cambridge University Press: Cambridge, UK, 2001; Volume 10, pp. 126–132.

9. Wallace, M.J. *Teaching Vocabulary*; Heinemann Educational Book: London, UK, 1982; p. 30.

10. Nation, I.S.P. *Teaching and Learning Vocabulary*; Foreign Language Teaching and Research Press: Beijing, China, 2004; p. 38.

11. Durrant, P.; Schmitt, N. To what extent do native and non-native writers make use of collocations? *Int. Rev. Appl. Linguist. Lang. Teach.* **2009**, *47*, 157–177. [CrossRef]

12. Nesselhauf, N. *Collocations in a Learner Corpus*; John Benjamins: Amsterdam, The Netherlands, 2005; pp. 1–344.

13. Farghal, M.; Obiedat, H. Collocations: A neglected variable in EFL. *IRAL-Int. Rev. Appl. Linguist. Lang. Teach.* **1995**, *33*, 315–331. [CrossRef]

14. McIntosh, C. *Oxford Collocations Dictionary for Student of English*; Oxford University Press: Oxfordshire, UK, 2009; pp. 456–480.

15. Koosha, M.; Jafarpour, A.A. Data-driven learning and teaching collocation of prepositions: The case of Iranian EFL adult learners. *Asian EFL J.* **2006**, *8*, 192–209.

16. Shokouhi, H.; Mirsalari, G.-A. Collocational knowledge versus general linguistic knowledge among Iranian EFL Learners. *TESL-EJ* **2010**, *13*, 21–24.

17. Zareie, A.; Koosha, M. Pattern of the Iranian Advanced Learners Problems with English Collocations: A Focus on Lexical Collocations. 2002. Available online: http://www.sid.ir/en/VEWSSID.J_pdf (accessed on 2 September 2010).

18. Boers, F.; Lindstromberg, S.; Eyckmans, J. Some explanations for the slow acquisition of L2 collocations. *Vigo Int. J. Appl. Linguist.* **2014**, *11*, 41–62.

19. Sinclair, J. *Corpus, Concordance, Collocation*; Oxford University Press: Cary, NC, USA, 1991.

20. Wolter, B.; Gyllstad, H. Frequency of input and L2 collocational processing: A comparison of congruent and incongruent collocations. *Stud. Second Lang. Acquis.* **2013**, *35*, 451–482. [CrossRef]

21. Yamashita, J.; Jiang, N. L1 influence on the acquisition of L2 collocations: Japanese ESL users and EFL learners acquiring English collocations. *Tesol Q.* **2010**, *44*, 647–668. [CrossRef]

22. Dang, T.N.Y.; Lu, C.; Webb, S. Incidental learning of single words and collocations through viewing an academic lecture. *Stud. Second Lang. Acquis.* **2022**, *44*, 708–736. [CrossRef]

23. Toomer, M.; Elgort, I. The development of implicit and explicit knowledge of collocations: A conceptual replication and extension of Sonbul and Schmitt (2013). *Lang. Learn.* **2019**, *69*, 405–439. [CrossRef]

24. Tsai, M.-H. The effects of explicit instruction on L2 learners' acquisition of verb–noun collocations. *Lang. Teach. Res.* **2020**, *24*, 138–162. [CrossRef]

25. Szudarski, P.; Carter, R. The role of input flood and input enhancement in EFL learners' acquisition of collocations. *Int. J. Appl. Linguist.* **2016**, *26*, 245–265. [CrossRef]

26. Alharbi, R.M.S. Acquisition of Lexical Collocations: A Corpus-Assisted Contrastive Analysis and Translation Approach. Ph.D. Thesis, Newcastle University, Newcastle, UK, 2017.

27. Boers, F.; Demecheleer, M.; Coxhead, A.; Webb, S. Gauging the effects of exercises on verb–noun collocations. *Lang. Teach. Res.* **2014**, *18*, 54–74. [CrossRef]

28. Snoder, P. Improving english learners' productive collocation knowledge: The effects of involvement load, spacing, and intentionality. *TESL Can. J.* **2017**, *34*, 140–164. [CrossRef]

29. Webb, S.; Kagimoto, E. The effects of vocabulary learning on collocation and meaning. *Tesol Q.* **2009**, *43*, 55–77. [CrossRef]

30. Webb, S.; Chang, A.C.S. How does mode of input affect the incidental learning of collocations? *Stud. Second Lang. Acquis.* **2022**, *44*, 35–56. [CrossRef]

31. El-Dakhs, D.A.S.; Ambreen, F.; Zaheer, M. The effect of textual enhancement on collocation learning: The case of Arab EFL learners. *Electron. J. Foreign Lang. Teach.* **2019**, *16*, 114–139.

32. Zhang, X.Y. Effects of receptive-productive integration tasks and prior knowledge of component words on L2 collocation development. *System* **2017**, *66*, 156–167. [CrossRef]

33. Cooker, L.; Torpey, M. From the classroom to the self-access centre: A chronicle of learner-centered curriculum development. *Lang. Teach.* **2004**, *28*, 11–16.

34. Chen, Y. The Effect of Learning Conditions on Collocation Gains: A Case Study of Task-based Dictionary Use Instruction. *Lexikos* **2022**, *32*, 1–30. [CrossRef]

35. Laufer, B. The contribution of dictionary use to the production and retention of collocations in a second language. *Int. J. Lexicogr.* **2011**, *24*, 29–49. [CrossRef]

36. Girgin, U. The Effectiveness of Using Corpus-Based Activities on the Learning of Some Phrasal-Prepositional Verbs. *Turk. Online J. Educ. Technol.-TOJET.* **2019**, *18*, 118–125.

37. Ganji, M. The effects on reading comprehension of lexical collocation instruction, subject matter knowledge, and cultural schema. *Theory Pract. Lang. Stud.* **2012**, *2*, 105–112. [CrossRef]

38. Johns, T. Should you be persuaded: Two samples of data-driven learning materials. *ELR J.* **1991**, *4*, 1–16.

39. Chan, T.P.; Liou, H.-C. Effects of Web-based Concordancing Instruction on EFL Students' Learning of Verb–Noun Collocations. *Comput. Assist. Lang. Learn.* **2005**, *18*, 231–251. [CrossRef]

40. Sun, Y.-C.; Wang, L.-Y. Concordancers in the EFL classroom: Cognitive approaches and collocation difficulty. *Comput. Assist. Lang. Learn.* **2003**, *16*, 83–94. [CrossRef]

41. Al-Mahbashi, A.; Noor, N.M.; Amir, Z. The effect of data driven learning on receptive vocabulary knowledge of Yemeni University learners. *3L Lang. Linguist. Lit.* **2015**, *21*, 13–24.

42. Römer, U. Corpus research applications in second language teaching. *Annu. Rev. Appl. Linguist.* **2011**, *31*, 205–225. [CrossRef]

43. Saeedakhtar, A.; Bagerin, M.; Abdi, R. The effect of hands-on and hands-off data-driven learning on low-intermediate learners' verb-preposition collocations. *System* **2020**, *91*, 102268. [CrossRef]

44. Boulton, A. Data-driven learning: On paper, in practice. *Corpus Linguist. Lang. Teach.* **2009**, *1*, 17–52.

45. Moher, D.; Liberati, A.; Tetzlaff, J.; Altman, D.G.; Prisma, G. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Int. J. Surg.* **2010**, *8*, 336–341. [CrossRef] [PubMed]

46. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Int. J. Surg.* **2021**, *88*, 105906. [CrossRef]

47. Sohrabi, C.; Franchi, T.; Mathew, G.; Kerwan, A.; Nicola, M.; Griffin, M.; Agha, M.; Agha, R. PRISMA 2020 statement: What's new and the importance of reporting guidelines. *Int. J. Surg.* **2021**, *88*, 105918. [CrossRef] [PubMed]

48. Fraenkel, J.R.; Wallen, W.E. *How to Design and Evaluate Educational Research*; McGraw Hill: New York, NY, USA, 2000; Volume 7.

49. Akbari, J.; Haghverdi, H.; Biria, R. Instructional efficacy of corpus-based tools in teaching collocations to Iranian university students with different majors. *J. Appl. Linguist. Lang. Res.* **2015**, *2*, 218–229.

50. Akinci, M.; Yildiz, S. Effectiveness of Corpus Consultation in Teaching Verb+Noun Collocations to Advanced ELT Students. *Eur. J. Appl. Linguist.* **2017**, *3*, 93–111. [CrossRef]

51. Almegren, A. Effect of Corpus-based Activities on Learning Verb–Noun Collocations in Saudi EFL Classes. *Jordan J. Mod. Lang. Lit.* **2022**, *14*, 371–383. [CrossRef]

52. Alruwaili, A.K. Data-driven Learning Approach for Teaching Verb-Noun Collocations in an English Foreign Language Context. *Asiatic* **2020**, *14*, 138–159.

53. Altun, H. The Learning Effect of Corpora on Strong and Weak Collocations: Impiicatiuns for Corpus-Basea Assessment of Collocation Competence. *Int. J. Assess. Tools Educ.* **2021**, *8*, 509–526. [CrossRef]

54. Ashouri, S.; Arjmandi, M.; Rahimi, R. The Impact of Corpus-Based Collocation Instruction on Iranian EFL Learners' Collocation Learning. *Univ. J. Educ. Res.* **2014**, *2*, 470–479. [CrossRef]

55. Daskalovska, N. Corpus-based versus traditional learning of collocations. *Comput. Assist. Lang. Learn.* **2015**, *28*, 130–144. [CrossRef]

56. DurmuŞ, B.Y.; KİLİMcİ, A. A corpus-driven deductive and inductive collocation instruction for young learners of English. *Int. J. Lang. Acad.* **2020**, *8*, 371–383.

57. Hirata, Y.; Hirata, Y. Applying 'Sketch Engine for Language Learning' in the Japanese English classroom. *J. Comput. High. Educ.* **2019**, *31*, 233–248. [CrossRef]

58. Huang, Z.P. The effects of paper-based DDL on the acquisition of lexico-grammatical patterns in L2 writing. *ReCALL* **2014**, *26*, 163–183. [CrossRef]

59. Jafarpour, A.A.; Hashemian, M.; Alipour, S. A corpus-based approach toward teaching collocation of synonyms. *Theory Pract. Lang. Stud.* **2013**, *3*, 51–60. [CrossRef]

60. Yunus, K.; Awab, S.A. The Effects of the Use of Module-based Concordance Materials and Data-driven Learning (DDL) Approach in Enhancing the Knowledge of Collocations of Prepositions among Malaysian Undergraduate Law Students. *Int. J. Learn.* **2012**, *18*, 181–198. [CrossRef]

61. Kim, H.; Chun, S. Fostering Lexis Awareness and Autonomy by Corpus-based Data-Driven Learning. *Engl. Teach.* **2008**, *63*, 213–235. [CrossRef]

62. Li, S. Using corpora to develop learners' collocational competence. *Lang. Learn. Technol.* **2017**, *21*, 153–171.

63. Łuszcz, M.A. Enhancing development of communicative skills in English through teaching English collocations in a corpus-based approacht. *Stud. About Lang.* **2016**, 88–101. [CrossRef]

64. Men, H.Y. Data-Driven Learning in Enhancing Learners' Language Idiomaticity. *Int. J. Emerg. Technol. Learn. (IJET)* **2020**, *15*, 27–41. [CrossRef]

65. Mohammadi Foomani, E.; Khalaji, K. Corpus-Based Versus Traditional Collocation Learning: The Case of Iranian EFL Learners. *J. Soc. Sci. Stud.* **2016**, *3*, 103–116. [CrossRef]

66. Oveidi, M.; Sepehri, M.; Shafiee, S. Differential Effects of Input/Output Tasks on Learning English Collocations by Iranian EFL Learners Through the Corpus-Based Instruction. *J. Mod. Res. Engl. Lang. Stud.* **2022**, *9*, 143–168. [CrossRef]

67. Rahimi, M.; Momeni, G. The effect of teaching collocations on English language proficiency. *Procedia-Soc. Behav. Sci.* **2012**, *31*, 37–42. [CrossRef]

68. Rezaee, A.A.; Marefat, H.; Saeedakhtar, A. Symmetrical and asymmetrical scaffolding of L2 collocations in the context of concordancing. *Comput. Assist. Lang. Learn.* **2015**, *28*, 532–549. [CrossRef]

69. Salehi, N.; Rasooyar, H. The effect of corpus-based collocation instruction on Iranian EFL learners' non-congruent collocation learning. *PONTE Int. Sci. Res. J.* **2016**, *72*, 33–51. [CrossRef]

70. Satake, Y. The effects of corpus use on learning L2 collocations. *JALT CALL J.* **2022**, *18*, 34–53. [CrossRef]

71. Uçar, S.; Yükselir, C. The effect of corpus-based activities on verb-noun collocations in EFL classes. *Turk. Online J. Educ. Technol.* **2015**, *14*, 195–205.

72. Wu, Y.-j.A. Discovering collocations via data-driven learning in L2 writing. *Lang. Learn. Technol.* **2021**, *25*, 192–214.

73. Yang, J. Effects of Collaborative Corpus-Based Learning on the Acquisition and Retention of Delexical Verb Effects of Collaborative Corpus-Based Learning on the Acquisition and Retention of Delexical Verb Collocation. *J. Stud. Lang.* **2015**, *31*, 67–94. [CrossRef]

74. Yilmaz, M. The Effect of Data-driven Learning on EFL Students' Acquisition of Lexico-grammatical Patterns in EFL Writing. *Eur. J. Appl. Linguist.* **2017**, *3*, 75–88. [CrossRef]

75. Luo, Q. *The Effects of Data-Driven Learning Activities on EFL Learners' Writing Development*; Oxford University Press: Oxfordshire, UK, 2016; Volume 5, pp. 1–13.

76. Coccetta, F. Multimodal Corpora and Concordancing in DDL. In *The Routledge Handbook of Corpora and English Language Teaching and Learning*; Routledge: London, UK, 2022; pp. 361–376. [CrossRef]

77. Cotos, E.; Link, S.; Huffman, S. Effects of DDL technology on genre learning. *Lang. Learn. Technol.* **2017**, *21*, 104–130.

78. Crosthwaite, P. *Data-Driven Learning for the Next Generation: Corpora and DDL for Pre-Tertiary Learners*; Routledge: London, UK, 2019.

79. Wang, Y.; Wu, L. College English classroom teaching design based on DDL. In Proceedings of the Fourth International Conference on Instrumentation and Measurement, Computer, Communication and Control, Harbin, China, 18 September 2014.

80. Hanafiyeh, M. The relationship between Iranian EFL learners 'multiple Intelligence and Success in Foreign Language Learning. *Asian J. Manag. Sci. Educ.* **2013**, *2*, 97–105.

81. Nesselhauf, N. The use of collocations by advanced learners of English and some implications for teaching. *Appl. Linguist.* **2003**, *24*, 223–242. [CrossRef]

82. Wray, A. *Formulaic Language and the Lexicon*; Cambridge University Press: Cambridge, UK, 2002.

83. Gao, Y.-M.; Zhang, Y. A Tentative corpus-based study of collocations acquisition by Chinese English language learners. *Can. Soc. Sci.* **2009**, *1*, 105–112.

84. Varley, S. I'll just look that up in the concordancer: Integrating corpus consultation into the language learning environment. *Comput. Assist. Lang. Learn.* **2009**, *22*, 133–152. [CrossRef]

85. Boulton, A. Beyond concordancing: Multiple affordances of corpora in university language degrees. *Procedia-Soc. Behav. Sci.* **2012**, *34*, 33–38. [CrossRef]

86. Chambers, A. Integrating corpus consultation in language studies. *Lang. Learn. Technol.* **2005**, *9*, 111–125.

87. Ashkan, L.; Seyyedrezaei, S.H. The effect of corpus-based language teaching on Iranian EFL learners' vocabulary learning and retention. *Int. J. Engl. Linguist.* **2016**, *6*, 190–196. [CrossRef]

88. Barabadi, E.; Khajavi, Y. The effect of data-driven approach to teaching vocabulary on Iranian students' learning of English vocabulary. *Cogent Educ.* **2017**, *4*, 1283876. [CrossRef]

89. Paker, T.; Ergül Özcan, Y. The effectiveness of using corpus-based materials in vocabulary teaching. *Int. J. Lang. Acad.* **2017**, *5*, 62–82. [CrossRef]

90. Schmidt, R.W. The role of consciousness in second language learning. *Appl. Linguist.* **1990**, *11*, 129–158. [CrossRef]

91. Rutherford, W.E.; Smith, M.S. *Grammar and Second Language Teaching: A Book of Readings*; Newbury House: New York, NY, USA, 1988.

92. Minaei, N.; Rezaie, G. The effects of collaborative and individual output tasks on learning English collocations. *J. Lang. Transl.* **2014**, *4*, 37–47.

93. Jahanbakhsh, A.A.; AliAsgariZamani, M.; Garman, Z. CIRC and STAD in Iranian context: Through the five elements to cooperative learning of lexical collocations. *Cogent Art Humanit.* **2019**, *6*, 1692469. [CrossRef]