



Corpus-Based Instruction for Developing Colligational Competence in Arab EFL Translation Students

Sahar Ali

Arab Open University, Egypt

Corresponding Author Email: saharxx66@yahoo.com

DOI: <https://doi.org/10.36892/ijlls.v8i5.2719>

APA Citation: Ali, S. (2026). Corpus-Based Instruction for Developing Colligational Competence in Arab EFL Translation Students. *International Journal of Language and Literary Studies*, 8(5).142-193. <https://doi.org/10.36892/ijlls.v8i5.2719>

Received:

04/06/2026

Accepted:

01/08/2026

Keywords:

Corpus-based instruction; data-driven learning (DDL); colligational competence; translation competence; Arab EFL learners

Abstract

The present research was conducted to examine the efficiency of corpus-based teaching in influencing the capacity of Arab English as a Foreign Language (EFL) university learners to recognize and use English colligational grammatical patterns in translation. A convergent mixed-methods design combined with a quasi-experimental one-group pre-test-post-test approach was used. Twenty-one first-year undergraduate translation students at the Arab Open University, Egypt, were involved in a 12-week Data-Driven Learning (DDL) intervention utilizing the British National Corpus (BNC), Sketch Engine, and BootCaT. Both quantitative and qualitative data were collected, the former through a translation achievement test administered before and after the intervention, and the latter through a post-intervention questionnaire and reflective audio recordings. Quantitative analyses involved descriptive statistics, the Kolmogorov-Smirnov test, paired-samples t-tests, and Cohen's d, while the qualitative data underwent thematic analysis. The results showed statistically significant changes in students' recognition and use of English colligational grammatical patterns, with both skills demonstrating large effect sizes. However, no major difference emerged in the case of longer translation activities, which implies that discourse-level translation competence needs a more sustained period of instructional support. Participants also reported improved grammatical awareness, increased confidence in making translation decisions, and greater reliance on authentic corpus evidence. These results affirm the need to incorporate corpus-based teaching into higher-education translation courses to strengthen grammatical knowledge and evidence-based translation practices among Arab EFL students.

1. Introduction

Translation is a complex cognitive and communicative process in which translators have to link linguistic, pragmatic and cultural knowledge to accurately and appropriately convey the meaning from one language to another (Hurtado Albir, 2017; Newmark, 2001; Venuti, 1995). In the field of higher education translation serves as central to the development of learners' bilingual competence, intercultural communication and professional skills (Kelly, 2005; Katan, 2004). Thus, translation competence has begun to be seen not as the ability to replace elements of source language with equivalents in the target language, but as the ability to make informed decisions about language choices that generate natural, coherent and contextually appropriate text in the target language (Hurtado Albir, 2017; Pym, 2010). To develop this level of competence, students must comprehend not only the patterns of words and grammar but also the patterns in which language is naturally organized (Nation, 2001; Bybee, 2006).

Another area of competence in translation is the knowledge of colligational grammatical patterns refers to the tendency of lexical items to occur in preferred grammatical environments

in which lexical items are used (Hoey, 2005; Hunston & Francis, 2000; Sinclair, 1991). Such patterns include verb–preposition combinations (e.g., depend on), adjective–preposition combinations (e.g., interested in), noun–preposition combinations (e.g., reason for), and verb complementation patterns (e.g., decide to do or avoid doing). Such grammatical associations are important to the overall fluency, accuracy and naturalness of translated texts because they represent conventional usage of the language rather than individual grammatical rules (Hunston & Francis, 2000; Sinclair, 2004a; Firth, 1957). Therefore, learners need to be able to grasp the relationship between lexical items and grammatical structures in real discourse in order to be successful in translation (Howarth, 1996; Hoey, 2005).

While colligational grammatical patterns are very important aspects to learn in the English language, they are among the most challenging aspects of English as a foreign language (EFL) (Nesselhauf, 2005; El-Dakhs, 2015; Boonraksa & Naisena, 2022). Learners can have a good amount of vocabulary and know each of the rules in grammar but have trouble matching grammatical structures with lexical items in a natural way. Because of this, such translations can be grammatically correct but at the same time be awkward, unnatural and not in line with the authentic use of English (Bahns, 1993; Nesselhauf, 2005; Mahmoud, 2005). When learners use the literal translation strategy or when learners transfer the grammatical structures directly from their first language (Odlin, 1989; Lado, 1957), these difficulties become more obvious.

This challenge is particularly evident among Arab EFL learners due to the significant differences between the syntactic structure, grammatical sequencing, complementation systems, and lexical-grammatical relations of Arabic and English languages (Farghal & Shunnaq, 1999; Mahmoud, 2005). Grammatical structures in Arabic are often quite unique and have no (direct) equivalents in English, while many English lexical items, in turn, need specific grammatical patterns in English that cannot be deduced from the meaning alone. As a result, negative language transfer may take place in the form of incorrect verb complementation, preposition usage and grammatical patterning while translating texts by Arab students (Odlin, 1989; Mahmoud, 2005; Mekahli, 2024). These frequent challenges impact not only the quality of the translation, but also the learners' capacity to write target texts that reflect authentic English language conventions (Metwally & Hamad, 2023; Keshavarz & Salimi, 2007).

The traditional approach to translation teaching in most EFL settings has focused mainly on memorizing words, explicit grammar teaching, and the teacher-centred instruction (Brown, 2006; Nation, 2009). These are all valuable ways that learners can gain basic language knowledge but they may also teach grammatical rules separate from their use in language. Typically, a translation exercise deals with finding the right grammatical structure rather than with the use of a lexical item in naturally occurring text. Thus, learners may rely primarily on intuition, dictionary definitions, or equivalents in their L1 in deciding how to translate, rather than from any authentic grammatical structure in the TL (Bowker, 2002; Newmark, 2001). Such teaching methods do not give pupils many chances to become aware of the lexical environments in which they are likely to find words, and in this way, they do not have many opportunities to make translation decisions that are appropriate to the context (El-Dakhs, 2015; Howarth, 1996).

The recent development of the field of corpus linguistics has created new opportunities for solving these problems by incorporating naturally occurring language data into language and translation learning (LTL) (McEnery & Hardie, 2012; McEnery et al., 2006; Biber et al., 1999). Language corpora are large, real language collections that contain naturally occurring written and spoken texts, which allow learners to see the way words and grammatical structures are used in real communicative situations. Rather than relying exclusively on prescriptive grammatical rules or learner intuition, corpus-based instruction enables learners to examine authentic language evidence when analysing recurrent grammatical patterns, comparing alternative linguistic choices, and verifying the appropriateness of translation decisions (Sinclair, 2004b; McEnery & Wilson, 2001). This approach bridges the gap between explicit

grammatical knowledge and authentic language use by exposing learners to naturally occurring language in context (Flowerdew, 2015; McEnery & Hardie, 2012).

In that context, DDL has become a significant approach for learners to study language in a learner-centered way, by using the corpus for language investigation (O'Keeffe & McCarthy, 2010; Johns, 1991). Rules are not provided by the teacher, but rather students are exposed to concordance lines, frequency data, collocational data, and contextual examples, in order to be able to recognize the common linguistic regularities. This exploration process encourages pupils to think of hypotheses, analyze the situation, and make decisions based on evidence, and to take more responsibility for their own learning (Boulton, 2010; Boulton & Vyatkina, 2021; Gilquin, 2021). Students receive support in making justifications for their choices of translation in translation classrooms, whereby DDL helps to enhance grammatical awareness, accuracy in translation, and learner autonomy in translation (Zanettin, 2009; Shawaqfeh & Khasawneh, 2023).

The use of corpus-based instruction has shown benefits for the acquisition of vocabulary, collocation, grammatical knowledge and writing skills as well as for autonomous learning in various EFL settings (Sun & Park, 2023; Uçar & Yükselir, 2015; Yılmaz & Soruç, 2015; Tsai, 2020). However, very little has been paid to the acquisition of English colligational grammatical patterns in translation education. Although the previous research has mostly focused on corpus use in general language learning, academic writing, or vocabulary learning, there are relatively few studies dealing with corpus consultation and learners' grammatical patterning recognition and application in their translation learning. (Vyatkina, 2016; Fang et al., 2021; Qiu, 2024) Moreover, although Arabic–English translation is considered one of the unique problems of translation in the Arabic context, empirical evidence is limited, particularly from Arab EFL contexts (Mekahli, 2024; Metwally & Hamad, 2023). As such, there are some important questions to be addressed about the degree to which English colligational patterns can be made salient to the learners and how such salientness can lead to better translations.

The present study addresses this gap by examining the effectiveness of corpus-based approach in enhancing the recognition and application of English colligational patterns in translation for Arab EFL university students. It is based on the principles of DDL, and the grammatical intervention combines authentic corpus resources with a well-structured translation program that invites students to explore the grammatical patterns, negotiate alternatives, and make decisions about translation supported by empirical evidence from the language (Johns, 1991; McEnery & Hardie, 2012; Sinclair, 2004b). In order to know whether the corpus-based method is effective in supporting the recognition and use of the English colligational grammatical patterns in translation, the study compares the performance of the learners in considering the congruence of the grammatical pattern before and after the instructional intervention.

The study contributes not only to translation pedagogy, but also to corpus-informed language instruction. It offers empirical evidence on the pedagogical usefulness of introducing the use of a corpus into university translation classes and practical advice for teachers who want to improve their students' grammatical competence using authentic language data (Boulton & Vyatkina, 2021; O'Keeffe & McCarthy, 2010; Zaki, 2021). Apart from that, the findings may also inform curriculum development by demonstrating how corpus-based instruction can complement traditional translation pedagogy to show that corpus-based teaching is used to supplement the traditional translation teaching method, encourage learners to have autonomy, and help learners to develop translation skills based on the results of active translation research (Canning, 2014; Timmis, 2015). The study focuses on Arab EFL learners, which is an under-researched aspect of L2 corpus studies on translation and also addresses the ongoing need for conducting targeted research on context-specific aspects of the use of technology in translation instruction (Hasumi & Chiu, 2024; Crosthwaite & Wijaya, 2023).

Although there is a large number of publications on Data-Driven Learning (DDL) and Corpus-based Translation pedagogy, there is still a gap between both of them in terms of both

empirical and theoretical aspects. In previous research, individual concordancers or general web corpora have been widely used by second language (L2) writers in correcting lexical errors but little interest has been shown in using several corpus resources simultaneously to correct highly specific grammatical patterns. Moreover, colligational competence, the structural and grammatical restrictions on the systematicity of word combinations under particular syntactic structures, has been known to be a major problem for undergraduate translators, especially in the context of Arab EFL learners who frequently face strong interference from the first language (L1). Very little research has been conducted so far that examines the way in which multiple corpus resources (parallel corpora, monolingual reference corpora, and self-made web-based corpora) are administered together to systematically develop this colligational competence in Arab EFL translation students. The current studies mainly concentrate on vocabulary learning or overall translation accuracy, with less focus on the cognitive and practical processes of multi-corpus navigation. This study comes in response to this, focusing on the effect of the blending of multi-corpus pedagogical approach on the acquisition of colligational patterns as well as the self-correction strategies of Arab undergraduate translation trainees.

Research Questions

The study seeks to answer the following research questions:

To what extent does corpus-based instruction improve Arab EFL university students' ability to recognize English colligational grammatical patterns in translation tasks?

To what extent does corpus-based instruction improve students' ability to apply English colligational grammatical patterns accurately during translation?

How do students perceive the use of corpus-based tools, including the BNC, Sketch Engine, and BootCaT, in supporting translation learning?

To what extent does corpus-based instruction influence students' overall translation performance?

Research Hypotheses

The study tests the following hypotheses:

H1. Corpus-based instruction significantly improves students' recognition of English colligational grammatical patterns.

H2. Corpus-based instruction significantly improves students' application of English colligational grammatical patterns during translation.

H3. Students demonstrate positive perceptions toward using corpus-based tools to support translation learning.

H4. Corpus-based instruction produces significant improvement in students' overall translation performance.

2. Literature Review

2.1 Corpus-Based Instruction in Translation Education

The introduction of corpus linguistics has significantly impacted on the world of language learning and teaching, as it now makes naturally occurring text data available for language teachers and learners. Corpus-based instruction allows learners to research language in use in different communicative settings rather than the traditional approach that focuses mainly on prescriptive grammatical rules and de-contextualized examples (Biber et al., 1999; McNery & Hardie, 2012; McNery et al., 2006). Learners can use systematic analysis of authentic linguistic evidence to identify patterns in the use of lexical and grammatical patterns, compare alternative expressions and assess the appropriateness of language choice from empirical evidence rather than intuition.

In translation education, the use of a corpus has gained significance as a tool in the learning process, since translation involves more than simply finding lexical equivalents. The professional translator is expected to create the target texts that are grammatically correct, contextually and naturally appropriate. Corpus resources can then help the learner confirm their

decisions, look at authentic contexts, contrast alternative linguistic forms and consider register variation and, finally, choose appropriate equivalents (Bowker, 2002; Zanettin, 2009; Kruger et al., 2011). Thus, teaching translation is based on the corpus, which promotes evidence-based decision making and helps to minimize the need for bilingual dictionaries or literal translation strategies.

The idea of corpus-based instruction is closely related to DDL introduced by Johns (1991) that is learning in which the learners construct linguistic knowledge themselves by directly interacting with corpus data. Learners do not simply 'read' grammatical rules into their head, but make inferences, read concordance lines, note regularities, and check their findings with authentic material. The inductive learning process fosters learner autonomy, analytical thinking, critical language awareness and enables students to be active learners (Boulton, 2010; Johns, 1991; O'Keeffe & McCarthy, 2010).

A number of empirical studies have proven the effectiveness of corpus-based instruction on the learners' acquisition of new vocabulary, grammatical accuracy, academic writing, collocation learning and strategies of independent language learning (Boulton & Vyatkina, 2021; Vyatkina, 2016; Yoon & Hirvela, 2004; Sun & Park, 2023). Also, the findings of the meta-analysis generally suggest that DDL has positive outcomes in various educational contexts provided that learners are given proper guidance in instruction (Boulton & Vyatkina, 2021; Gilquin, 2021). The results validate the use of corpus in language and translation teaching.

2.2 Colligational Competence in Translation

Colligational competence is one of the important parts of the translation competence that is defined as recognizing and producing the grammatical patterns that are often used with specific lexical units. Collocation is about what lexical items go with each other as in collocation. Colligation refers to the bond between lexical items and their preferred grammatical environment (Firth, 1957; Hoey, 2005; Hunston & Francis, 2000). These preferences are related to grammatical structures such as verb–preposition combinations, adjective–preposition combinations, noun complementation, verb complementation patterns, and others that occur frequently in authentic language use.

Lexical equivalence is not enough for natural production in the target language, hence the need for colligational competence in translation. The translator needs to choose the grammatical structures which are close to the norms in the target language. The grammatical preferences are frequently ignored, and this leads to non-pragmatic translations that can be grammatically correct, but stylistically unnatural (Hunston & Francis, 2000; Sinclair, 1991; Hoey, 2005).

Nesselhauf (2005), Mahmoud (2005) and El-Dakhs (2015) conducted several studies that consistently found that EFL learners find English colligational patterns very difficult to grasp, especially when their L1 colligational patterns differ from English (Nesselhauf, 2005; Mahmoud, 2005; El-Dakhs, 2015). In the case of the Arab learners, these challenges are exacerbated by significant variations between the Arabic and English syntax, as well as complementation and the use of prepositions (Farghal & Shunnaq, 1999; Mahmoud, 2005; Mekahli, 2024). As a result, students often use Arabic grammatical constructions in English, which leads to frequent errors in verb complementation, adjective complementation and preposition use (Odlin, 1989; Lado, 1957).

Common errors are depend from instead of depend on, interested by instead of interested in, reason of instead of reason for, and participate to instead of participate in. The errors show language transfer from negative and highlight how learners often resort to literal rather than to the English conventions. These results bring about the need to adopt pedagogical methods that bring the learner into contact with the grammatical patterns that exist in the language (Keshavarz & Salimi, 2007; Metwally & Hamad, 2023).

2.3 Corpus Tools Supporting Translation Learning

Good corpus-based instruction requires corpus resources which can help the students investigate the authentic language in systematic ways. This present study uses three linguistic tools: corpus tools, which are complementary and all offer different types of linguistic evidence to support translation learning.

The BNC is a balanced spoken and written corpus of British English which covers various registers and communication contexts (Burnard, 2007). The BNC provides the learner with authentic language used by native speakers, allowing them to see naturally occurring patterns of grammar and learn about contextual variation, and to check on their translations with linguistic evidence that can be trusted.

Other features of Sketch Engine include advanced analytical capabilities such as concordance search, Word Sketch analysis, collocation extraction, grammatical pattern identification and frequency profiling (Kilgarriff et al., 2014). These tools enable students to explore the lexicological behaviour in a systematic way and to compare grammatical environments in vast amounts of authentic language data.

The BootCaT Toolkit is a tool that allows users to generate custom corpora from real web-based text with the help of keywords (Baroni et al., 2004). A major advantage of specialized corpora is that they provide learners with authentic words, phrases and structures from particular subject areas like law, medicine, business or technology, etc. Combined these three resources provide a full-fledged corpus-learning system for general language learning and for specialized translation training.

2.4 Recent empirical Studies on Corpus-Based Translation Learning

Pedagogical studies on the effectiveness of using a corpus in language and translation teaching have been growing. The results of the studies consistently show that students learning with corpus-based instruction make better progress in the following areas than those with traditional instruction: grammatical accuracy, vocabulary development, collocational knowledge, writing ability, and learner autonomy (Boulton & Vyatkina, 2021; Vyatkina, 2016; Yoon & Hirvela, 2004; Sun & Park, 2023). The availability of evidence from the corpus also prompts students to support their decisions with authentic evidence, thus fostering analytical thinking and minimizing the use of intuition (Flowerdew, 2015; Gilquin, 2021).

In the field of translation education, corpus-driven teaching has also been shown to enhance lexical choice, the accuracy of grammar, contextual relevance, and the quality of translated texts (Bowker, 2002; Zanettin, 2009; Fang et al., 2021). Consultation of authentic corpora typically leads to learners' improved abilities to assess competing translation options, to identify textual conventions based on grammar and to produce text in the target language that is more similar to native speaker use.

However, there are also some difficulties involved in the implementation of a corpus. An initial problem that novice users have with the use of concordance is how to interpret lines in the concordance (Boulton, 2010; Canning, 2014); another is how to select an appropriate search strategy (Boulton, 2010), and another is the understanding of the output of the corpus, especially early in the learning process (Canning, 2014; Pasquale, 2018). Therefore, even though the learning materials are available in the target language, the learners benefit from full integration from the corpus only through explicit training, guided practice, scaffolding by the teacher, and design of learning tasks (Timmis, 2015; Crosthwaite & Wijaya, 2023).

Although corpus-based method has already proven to be effective, few studies have investigated the effect of corpus-based method on the development of English colligational competence in translation courses for university students of Arabic as a foreign language (EFL). While previous studies have focused on vocabulary learning, collocation acquisition, writing instruction, or general grammar development, the acquisition of colligational grammatical patterns during translation and their application have been relatively under-researched areas (Vyatkina, 2016; Fang et al., 2021; Qiu, 2024). In addition, little work has been focused on the integration of several corpus tools in the structured programme of

instruction and the translation performance and perceptions of learners (Shawaqfeh & Khasawneh, 2023; Hasumi & Chiu, 2024). The limitations have been the bases for the present study aimed at enriching the existing literature by testing the effectiveness of corpus-based approach in improving Arab EFL learners' recognition and use of English colligational patterns in translation.

Collectively, the literature reviewed confirms that, in the field of L2 acquisition, corpus-based approach has a positive effect on several L2 domains such as vocabulary, grammar, collocational competence, quality of translation, and learner autonomy (Boulton & Vyatkina, 2021; Sun & Park, 2023; Fang et al., 2021). There is, however, little empirical research to test its impact on English colligational grammatical patterns in the translation of Arabic learners into English, employing various corpus tools in a well-designed instructional programme. This study was aimed to fill this gap by employing a mixed methods quasi-experimental design. Therefore, the methods used in this study are described in the following section to investigate the effectiveness of corpus-based teaching in developing the recognition and use of English colligational patterns in the translation process of Arab EFL university students.

Empirical study of the practical application and efficiency of DDL in a variety of translation and L2 writing tasks has been more and more concentrated in recent years. For example, Tsai (2020) showed that corpus-assisted translation teaching has a significant positive effect on students' lexical selection, and that the students' awareness of stylistic variation is raised. Recent studies by Crosthwaite and Wijaya (2023) highlight that teaching corpus literacy will not only enable the student to correct his/her own errors in real time, but also provide long-term support in terms of establishing learner autonomy. Shawaqfeh and Khasawneh (2023) noted that the use of corpus queries in the Arab EFL context could be effective in reducing the persistent L1 collocational interference and making abstract translation rules concrete lexical decisions. Moreover, recent research has started to look into the integration of the traditional corpus tools with the new educational technologies. Hasumi and Chiu (2024) examined the student's transition from a monolingual corpus towards a bilingual concordancer and found that multi-tool corpus environments enable students to engage in more complex thinking. Likewise, Qiu (2024) investigated the synergy of corpus-based revision and generative AI tools and found that corpus-verified proofing is an important pedagogical measure against errors and hallucinations sometimes created by AI translation engines. These studies are a testament to the clear paradigm shift that has taken place in the field of translator training: today, the training process needs to provide students with a multi-corpus toolkit so that they can acquire strong and digitally equipped translation skills.

3. Methodology

3.1 Research Design

This study employed a mixed-methods convergent design to investigate the effectiveness of corpus-based instruction in developing first-year university students' recognition and application of English colligational grammatical patterns in translation. A mixed-methods approach integrates quantitative and qualitative data to provide a more comprehensive understanding of educational phenomena than either method alone (Creswell & Creswell, 2018). Quantitative data were collected to measure changes in students' performance following the instructional intervention, whereas qualitative data were gathered to explore learners' experiences and perceptions of corpus-assisted translation learning.

Because the participants were enrolled in a single intact translation class, a quasi-experimental one-group pre-test–post-test design was adopted. Students completed a translation achievement test before and after the instructional intervention, allowing changes in their performance to be examined over time. To strengthen the credibility and interpretation of the findings, the quantitative results were triangulated with data obtained from the post-intervention questionnaire and students' reflective audio recordings (Creswell & Creswell, 2018).

3.2 Research Context and Participants

The study was conducted during the first semester of the academic year at the Arab Open University (AOU), Egypt, within a compulsory translation course for first-year students enrolled in the English Language Programme. Before the intervention, participants had completed general English language courses but had received no formal instruction in corpus linguistics, Data-Driven Learning (DDL), or corpus-assisted translation (CAT).

Convenience sampling was used since the participants were chosen from the teaching situation the researcher had access to. Firstly, first-year undergraduate students who took the compulsory translation course were recruited ($n = 33$). Nevertheless, the number of students who went through all study phases and completed the final analysis was limited to 21 ($N = 21$). The other 12 students dropped out due to time constraints, higher workloads, technical issues, or personal commitments, and not because of the instructional intervention. Students who came late frequently or did not provide thorough assessment data were excluded from the final analysis in order to maintain the accuracy and completeness of the dataset.

The final sample consisted of 21 first-year undergraduate students, all of whom were native speakers of Arabic in the English Language Programme. Their level of English ranged from beginner to intermediate, with none having prior formal training in corpus linguistics, Data-Driven Learning (DDL), or corpus-assisted translation (CAT). During the 12-week intervention, participants were given structured corpus-based learning activities, attended all the instructional sessions, sat the pre-test and the post-test, answered the post-intervention questionnaire, and were asked to take part in qualitative reflections using audio recordings.

The use of the convenience sampling technique was adopted in this study because of institutional parameters and curricular constraints. The sample consisted of the 21 undergraduate students in a pre-existing compulsory translation course. Convenience sampling does not allow for any broad generalizability of the results, but the method was very suitable as it enabled the seamless and ethical incorporation of the corpus training framework directly in the participants' active credit-bearing coursework. The context of the natural classroom reduced the influence of external ecological factors, and the teaching and learning activities were directly related to the short-term academic needs and learning requirements of students.

Table 1. Participants' Profile

Variable	Category	Frequency	Percentage
Participants	Completed the study	21	100%
	Withdrawals	12	36.4% of recruited sample
Native Language	Arabic	21	100%
Academic Level	First year	21	100%

3.3 Instructional Intervention

The instructional programme was carried out over 12 weeks, with a total of 42 instructional hours made up of 9 hours of preparation, 27 hours of guided instruction, and 6 hours of assessment. The intervention was developed based on the principles of Data-Driven Learning (DDL), which promotes learners discovering linguistic regularities by actively exploring authentic corpus data instead of depending purely on teacher-led explanation (Johns, 1991; Boulton & Vyatkina, 2021; O'Keeffe & McCarthy, 2010).

Instruction focused on the ability to identify and apply the colligational grammatical patterns commonly found in translation, especially verb–preposition, adjective–preposition, noun–preposition, and verb complementation patterns (Hunston & Francis, 2000; Hoey, 2005; Sinclair, 2004b). As the programme went on, learners moved from highly guided corpus consultation to increasingly independent corpus analysis and application when performing translation tasks, becoming more confident and autonomous in corpus-assisted translation (McEnery & Hardie, 2012; Timmis, 2015).

The training programme was structured into three sequential phases, as outlined in Table 2.

Table 2- Summary of the Instructional Phases

Phase	Weeks	Instructional Focus
Phase 1	1–4	Introduction to corpus linguistics, Data-Driven Learning (DDL), and training in corpus tools
Phase 2	5–10	Guided corpus consultation and application of English colligational grammatical patterns in translation tasks
Phase 3	11–12	Post-test administration, questionnaire completion, and collection of qualitative reflections

3.4 Instructional Procedures

In the training stage, students were taught how to search corpus databases, process concordance lines, recognize grammatical contexts, and analyse natural language patterns (Sinclair, 2004b; McEnery & Hardie, 2012). The application phase then incorporated corpus consultation into both sentence-level and short-text translation assignments, which allowed learners to contrast different grammatical forms, assess linguistic options, and defend their translation choices with authentic corpus evidence (Bowker, 2002; Zanettin, 2009). Post-intervention assessment and qualitative data collection made up the last phase.

Training was based on a six-stage workflow inspired by corpus-based pedagogical practices (Sinclair, 2004b; McEnery & Hardie, 2012; Timmis, 2015):

- Corpus exploration
- Concordance analysis
- Identification of grammatical patterns
- Data-Driven Learning (DDL) activities
- Corpus-assisted translation practice
- Reflection and instructor feedback

The scaffolding of instruction was provided during the intervention through live demonstrations, guided practice, annotated examples, instructional manuals, and ongoing instructor feedback. Such support has been broadly advised as a way of helping beginner corpus users successfully interpret concordance data and gradually build learner autonomy (Boulton & Vyatkina, 2021; Canning, 2014; Timmis, 2015; Vyatkina, 2016).

3.5 Corpus Resources

The instructional programme incorporated three complementary corpus resources.

The main reference corpus was the British National Corpus (BNC), which offered genuine examples of both spoken and written British English across diverse genres (Burnard, 2007; McEnery & Hardie, 2012). The corpus was consulted so that students could explore natural grammatical usage and confirm their translation decisions.

Sketch Engine helped to provide more sophisticated linguistic analysis through concordance searches, Word Sketch, collocation extraction, grammatical pattern detection, and frequency profiling (Kilgarriff et al., 2014). These functions allowed learners to explore associations between lexical items and their favoured grammatical contexts.

The BootCaT Toolkit allowed students to build discipline-specific corpora out of real-life online texts with the help of discipline-specific keywords (Baroni et al., 2004). These corpora introduced learners to authentic terminology and grammatical conventions associated with particular fields of translation. Combined, these resources were complementary in the investigation of authentic language use, the comparison of grammatical options, and evidence-based translation decision-making (Bowker, 2002; McEnery & Hardie, 2012; Zanettin, 2009).

3.6 Research Instruments

To collect both quantitative and qualitative data, three research instruments were used.

The translation achievement test was given before the instructional intervention and after it to determine students' ability to recognize and use English colligational grammatical patterns in translation. The test included three parts: (a) identification of grammatical patterns, (b) usage of colligational grammatical patterns, and (c) translation of sentences and short texts. The test was created on the basis of the translation competence literature and corpus-based language learning (Hurtado Albir, 2017; Hoey, 2005; McEnery & Hardie, 2012). The translation component scores were measured using an analytic scoring rubric that dealt with grammatical correctness, situational suitability, and the correct use of English colligational grammatical patterns.

The post-intervention questionnaire was given to assess learners' perceptions of the usability of the corpus tool, the effectiveness of the instruction, learner grammatical awareness, learner autonomy, and the general learning experience. The questionnaire was developed based on previous research on corpus-based language learning and Data-Driven Learning (Boulton & Vyatkina, 2021; Vyatkina, 2016; Yoon & Hirvela, 2004). It involved 21 Likert-scale questions along with some open-ended questions, which would allow gathering qualitative feedback on the advantages and problems of corpus-assisted translation learning.

Qualitative data were collected through students' reflective audio recordings (WhatsApp voice messages). Participants described their experiences using the corpus tools, their perceived learning gains, the challenges they encountered, and the influence of corpus consultation on their translation decisions. All audio recordings were transcribed verbatim, anonymized using participant codes, and translated into English where necessary.

3.7 Validity and Reliability of the Research Instruments

In order to ascertain the quality of the research instruments, appropriate procedures were followed to determine their reliability and validity. The translation achievement test and the post-intervention questionnaire were developed based on the literature on corpus-based language learning, Data-Driven Learning, and translation pedagogy (Hurtado Albir, 2017; Johns, 1991; Kelly, 2005; McEnery & Hardie, 2012). The content validity of both instruments was established by having them reviewed by experts in translation studies, corpus linguistics, applied linguistics, and English language teaching. The relevance, clarity, organisation, and how well the instruments matched the study objectives were assessed by the reviewers. The final versions of the instruments incorporated their comments and suggestions prior to the main study (Creswell & Creswell, 2018).

Cronbach's alpha was used to determine the internal consistency of the post-intervention questionnaire. As shown in Table 3, the questionnaire yielded a Cronbach's alpha coefficient of .839, which is high and indicates good reliability, showing that the instrument was consistent in measuring participants' perceptions of corpus-based instruction.

Table 3-Reliability of the Questionnaire

Instrument	No. of Items	Cronbach's α
Questionnaire evaluating the use of the BNC, Sketch Engine, and BootCaT	21	.839

3.8 Data Analysis

Before performing inferential statistical analysis, normality was tested using the Kolmogorov–Smirnov test. The findings revealed that the pattern identification ($p = .090$), pattern application ($p = .120$), and translation task ($p = .520$) distributions did not significantly differ from a normal distribution. Based on this, paired-samples t tests were deemed suitable for comparing participants' pre-test and post-test results.

Quantitative data were analysed using IBM SPSS Statistics 29. Descriptive statistics (Ms and SDs) were calculated to summarize participants' performance on the pre-test and post-test measures. Paired-samples t tests were conducted to determine whether statistically significant differences existed between pre-test and post-test scores for pattern identification, pattern

application, translation performance, and overall achievement. To assess the magnitude of the instructional effects, Cohen's *d* was calculated and interpreted according to the conventional benchmarks proposed by Jacob Cohen (1988).

Qualitative data obtained from the post-intervention questionnaire and participants' reflective audio recordings were analysed using thematic analysis following the procedures proposed by Braun and Clarke (2006). The transcripts were read repeatedly to generate initial descriptive codes, which were subsequently organized into broader themes reflecting learners' perceptions of corpus-based instruction. The final themes focused on the perceived usefulness of corpus tools, the development of grammatical awareness, learner autonomy, confidence in translation decision-making, and challenges encountered during corpus consultation.

Finally, the quantitative and qualitative findings were integrated during the interpretation stage to provide a comprehensive evaluation of the effectiveness of corpus-based instruction in enhancing students' recognition and application of English colligational grammatical patterns in translation, consistent with the principles of mixed-methods research (Creswell & Creswell, 2018).

3.9 Ethical Considerations

Consent to participate in the study was freely given and informed consent was obtained from each participant before data collection. The aim of the study, the procedures and the right of the student to stop the study at any time without any loss of academic credit were explained to the students. Participant codes were used to anonymize all responses to ensure confidentiality. Audio reflections were transcribed and safely archived for research purposes only. The study abided by the ethical conduct of the Arab Open University.

4. Results

This section presents the quantitative findings, followed by the questionnaire and qualitative findings, to evaluate the effectiveness of the corpus-based instructional intervention.

4.1 Quantitative Findings

Pre- test and Post test Performance

The participants' performance was compared prior to and following the 12 - week corpus-based instructional program using descriptive statistics. The mean scores and standard deviations for each component of the achievement test are summarized in Table 4.

Table 4 - Pre-test and Post-test Achievement scores

Assessment Component	Pre-test M (SD)	Post-test M (SD)	t	p	Cohen's d
Pattern Identification	16.43 (4.68)	20.90 (4.28)	4.438	.001	0.96
Pattern Application	23.38 (7.65)	29.43 (5.71)	3.735	.001	0.82
Translation Tasks	10.29 (4.31)	10.10 (3.43)	0.499	.620	0.05
Overall Achievement	50.10 (11.79)	60.43 (10.86)	4.514	.001	0.98

Note. M = mean; SD = standard deviation. Statistical significance was set at $p \leq .05$.

4.2 Assessment Components

Participants showed higher mean scores on pattern identification, pattern application, and overall achievement after the intervention. Specifically, participants' mean score for pattern identification increased from 16.43 (SD = 4.68) to 20.90 (SD = 4.28), while the mean score for pattern application increased from 23.38 (SD = 7.65) to 29.43 (SD = 5.71). Overall achievement also improved substantially, rising from 50.10 (SD = 11.79) to 60.43 (SD = 10.86).

Paired-samples t-tests confirmed that these gains were statistically significant for pattern identification ($t = 4.438$, $p = .001$, $d = 0.96$), pattern application ($t = 3.735$, $p = .001$, $d = 0.82$), and overall achievement ($t = 4.514$, $p = .001$, $d = 0.98$), with large effect sizes indicating substantial instructional impact. In contrast, performance on the extended translation task showed no statistically significant improvement ($t = 0.499$, $p = .620$, $d = 0.05$), suggesting that although corpus-based instruction effectively enhanced learners' recognition and application of English colligational grammatical patterns, these gains did not extend to performance on longer translation tasks within the duration of the intervention.

4.3 Effect Size of the Instructional Intervention

To assess the practical significance of the instructional intervention, Cohen's d was calculated for each assessment component. The results are presented in Table 5.

Table 5. Use of Corpus Tools

Assessment Component	Cohen's d	Interpretation
Pattern Identification	0.96	Large
Pattern Application	0.82	Large
Translation Task	0.05	Negligible
Overall Achievement	0.98	Large

The effect size analysis indicates that the intervention had a large practical impact on students' ability to identify and apply English colligational grammatical patterns, as well as on their overall achievement. In contrast, the effect on the extended translation task was negligible ($d = 0.05$), suggesting that although corpus-based instruction substantially improved sentence-level grammatical competence, these gains did not readily transfer to more cognitively demanding translation tasks within the twelve-week intervention period.

4.4 Questionnaire Findings

The post-intervention questionnaire examined participants' perceptions of corpus-based instruction, patterns of corpus tool use, and the perceived impact of the instructional programme on the development of English colligational grammatical competence in translation.

Corpus Tool Usage

Participants reported frequent use of the corpus tools throughout the instructional programme. Table 6 summarizes the corpus tools most frequently used by the participants.

Table 6. Participants' Use of Corpus Tools During the Instructional Programme

Corpus Tool	Frequency	Percentage
Sketch Engine	10	47.6
BNC	9	42.9
BootCaT	2	9.5
Total	21	100.0

Sketch Engine was the most frequently used corpus resource (47.6%), followed by the BNC (42.9%), whereas BootCaT was used less frequently (9.5%).

Notably, BootCaT had the lowest frequency of use of any of the tools assessed (9.5% of participants used it regularly). Two main reasons for this low uptake were identified from the qualitative follow-up. Firstly BootCaT is very specialized in the construction of domain specific corpora from the web, which is more of a technical learning curve than is the case for searching pre-compiled resources such as Sketch Engine or BNC. Third, the translation process for this study focussed mainly on the errors made at the sentence level and not so much on highly specialized technical words, so that students preferred more direct and ready-to-use concordancers rather than having to create their own web corpora through a multi-step process, especially in the limited time frame of classroom sessions.

The corpus functions most frequently performed by the participants are presented in Table 7.

Table 7 - Frequency of Corpus Tool Use

Response	Frequency	Percentage
Rarely	2	9.5
Sometimes	8	38.1
Every session	11	52.4
Total	21	100.0

Participants reported regular use of the corpus tools throughout the instructional programme. As shown in Table 8, over half of the participants (52.4%) indicated that they used the corpus tools during every instructional session, while 38.1% reported using them sometimes. Only 9.5% reported using the tools rarely. These findings suggest a high level of engagement with the corpus resources during the instructional intervention.

Participants were also asked whether they encountered any difficulties while using the corpus tools. As shown in Table 9, most participants (81.0%) reported experiencing no major difficulties during corpus consultation, whereas only 19.0% indicated that they encountered some challenges.

Table 8. Difficulties Encountered During Corpus Use

Response	Frequency	Percentage
No	17	81.0
Yes	4	19.0
Total	21	100.0

All participants (100%) indicated that the instructional guidance provided throughout the programme was clear and easy to follow (Table 9).

Table 9. Perceived Clarity of Instruction

Response	Frequency	Percentage
Yes	21	100.0
No	0	0.0
Total	21	100.0

Perceived Skill Development

Participants evaluated the extent to which the corpus-based instructional programme contributed to the development of their translation-related skills. Overall, learners reported positive perceptions of the instructional programme, with a mean score of 4.14 (SD = 0.31), corresponding to a relative importance of 82.86%. These findings indicate that participants generally perceived the corpus-based instructional programme as effective in enhancing their translation-related skills. The highest-rated questionnaire items are presented in Table 10.

Table 10 - Highest-Rated Questionnaire Items

Item	Mean	Relative Importance (%)
Awareness of English colligational grammatical patterns	4.48	89.52
Confidence in grammatical decision-making	4.43	88.57
Translation decision-making	4.29	85.71

As shown in Table 10, participants rated increased awareness of English colligational grammatical patterns most highly (M = 4.48, Relative Importance = 89.52%), followed by confidence in grammatical decision-making (M = 4.43, Relative Importance = 88.57%) and translation decision-making (M = 4.29, Relative Importance = 85.71%). The lowest-rated item concerned independent corpus use (M = 3.71, Relative Importance = 74.29%), although responses remained within the positive range, suggesting that participants perceived corpus consultation positively while still requiring additional support to develop independent corpus-use skills.

4.5 Qualitative Findings

Qualitative data obtained from participants' reflective audio recordings complemented the questionnaire findings by providing additional insights into learners' experiences with corpus-based instruction.

Thematic analysis identified four recurrent themes:

Improved awareness of English colligational grammatical patterns.

Increased confidence in translation decision-making.

Greater learner autonomy through corpus consultation.

Initial technical challenges in using corpus tools.

In many instances, participants also said that using authentic corpus data allowed them to tell the difference between grammatically correct phrases and their natural use in English. Some of the respondents mentioned that they replaced literal translations like *afraid from* with *afraid of* and *do an effort with* *make an effort* after referring to concordance lines and collocational data in Sketch Engine and the BNC.

This shift from intuitive, literal translation to corpus-based verification of evidence is very noticeable in the students' self-reports. The qualitative interviews often revealed that corpus data helped to clarify fixed literal translations. For instance, Participant 3 said:

When I used to do a lot of translating from English into Arabic, I would always translate 'afraid from' from 'من خائف' - but when I did a quick search on BNC, I saw this wall of 'afraid of'. It was an eye opener.

Likewise, Participant 12 noted that corpus queries helped to address problems with light verb collocations:

I've always preferred to use the word, 'do', in my essays, because in my head it was 'يفعل', which means 'make' in English, and when I checked the word sketch of Sketch Engine, I saw that, indeed, it was 'make'. So I have to look up my verbs before I put them in my essay.

This dynamic cognitive shift was also referred to as an increase in general writing confidence. Participant 7 remarked:

Before this training I was just guessing and hoping that a word translates as I thought it translated; now when I am unsure about the word that I'm translating I can use the corpus to find out if real native writers use it that way; it gives me a logical reason to choose between which translation is which.

Last, Participant 15 explained the pedagogical benefit of DDL compared to conventional dictionaries:

The corpus search revealed me the precise grammatical situation of the words I wanted to use, whereas standard dictionaries just tell me what the word means and don't give me the surrounding grammar and prepositions.

Another finding was that participants relied less on teacher explanations and bilingual dictionaries toward the end of the instructional programme, instead using corpus evidence to justify their translation decisions. Participants noted some initial challenges with understanding concordance lines and navigating corpus interfaces, but these problems decreased with guided practice and the help of the instructor.

In response, the most common suggestions given by the participants were to add more practical corpus-based activities and to extend the length of the instructional programme, giving them more time to consult the corpus on their own.

5. Discussion

To evaluate the wider scholarly value of these findings, Table 11 systematically contrasts the results of this study with prominent prior empirical literature in the field of DDL and translator training.

Table 11- Comparative Summary of Current Findings with Prior Empirical Literature

Study	Research Context	Key Findings	Points of Alignment with Present Study	Unique Contributions of Present Study
Yoon & Hirvela (2004)	ESL Writers (USA)	Corpora improve L2 writing confidence and grammatical accuracy.	Confirms that corpus tools significantly reduce grammatical errors and build student autonomy.	Evaluates a multi-corpus model specifically targeting translation colligational

				competence rather than general L2 writing.
Uçar & Yükselir (2015)	EFL Learners (Turkey)	DDL is highly effective for pre-intermediate learners mastering prepositions.	Agrees that prepositions and structural collocations are the areas of greatest immediate student improvement.	Demonstrates these gains within an advanced undergraduate translation curriculum rather than general English language classes.
Crosthwaite & Wijaya (2023)	L2 Writers (Asia)	Long-term corpus training leads to sustained improvements in L2 error self-correction.	Supports the finding that hands-on corpus training shifts students from passive learners to active researchers.	Illustrates the specific technical hurdles (e.g., BootCaT's learning curve) in short-term curricular frameworks.
Present Study (2026)	Arab EFL Translation Undergraduates	Multi-corpus training significantly improves colligational competence and reduces L1 interference.	Aligns with prior research on the viability of DDL for self-correction and reducing lexical-grammatical errors.	Focuses specifically on Arab EFL translation trainees; provides empirical evidence on the under-utilization of custom web-corpora (BootCaT) due to task-specific utility and cognitive load.

5.1 Interpretation of Quantitative and Qualitative Findings

This paper explored how effective corpus-based instruction is in training Arab EFL university students to recognize and use English colligational grammatical patterns in translation. The study combined the BNC, Sketch Engine, and BootCaT within a DDL framework to examine both the measurable learning outcomes and the perceptions learners had of corpus-assisted translation instruction. The mixed-methods results supplement the evidence for the idea that corpus-informed pedagogy can significantly improve learners' grammatical awareness and sentence-level translation performance while also fostering learner autonomy and evidence-based decision-making (Johns, 1991; McEnery & Hardie, 2012; Boulton & Vyatkins, 2021).

The quantitative results show that students significantly increased their skill in identifying and using English colligational grammatical patterns after the twelve-week instructional intervention. Pattern identification ($d = 0.96$), pattern application ($d = 0.82$), and overall achievement ($d = 0.98$) have large effect sizes, demonstrating that the intervention produced meaningful educational gains as opposed to merely statistically significant differences. These results confirm the argument that genuine corpus consultation helps learners see recurrent grammatical environments and make translation choices on the basis of empirical linguistic evidence rather than intuition or literal transfer from the first language (Sinclair, 2004b; Hoey, 2005; McEnery & Hardie, 2012).

These results align with other studies on DDL indicating that corpus consultation facilitates grammatical awareness and vocabulary growth, as well as learner autonomy, by encouraging students to research authentic language on their own (Johns, 1991; Boulton, 2010; Boulton & Vyatkins, 2021; Gilquin, 2021). Instead of memorizing isolated grammatical rules, learners explored concordance lines, contrasted authentic examples, recognized recurring grammatical patterns, and checked the validity of their translation decisions with corpus evidence. This kind of inductive learning reflects the central concepts of DDL and explains the substantial gains observed in learners' recognition and application of colligational grammatical patterns.

The findings further support corpus-linguistic approaches to colligation, which argue that grammatical behaviour cannot be isolated from lexical meaning and that genuine language use reveals observable, predictable grammatical preferences that cannot be learnt through

conventional grammar instruction alone (Firth, 1957; Hunston & Francis, 2000; Hoey, 2005; Sinclair, 1991). Corpus consultation appears to reduce negative language transfer in Arab EFL learners, since their first language differs considerably from English in syntactic structure, complementation, and use of prepositions (Farghal & Shunnaq, 1999; Mahmoud, 2005; Mekahli, 2024). This process is illustrated by the qualitative results, with participants citing literal translations such as afraid from and do an effort and changing them into more natural expressions like afraid of and make an effort after consulting the corpus evidence.

5.2 Micro- vs. Macro-level Translation Competence

Despite the high level of improvement in grammatical pattern recognition and usage, no statistically significant improvement with respect to the extended translation task was observed. This result indicates that any gain in micro-level grammatical competence may not necessarily carry over to macro-level performance in translation over a relatively brief period of instruction. Extended translation involves the coordination of various mental activities and processes, such as discourse structure, textual unity, stylistic suitability, lexical choice, pragmatic interpretation, and audience awareness. Therefore, although corpus consultation successfully enhanced sentence-level grammatical decision-making, learners might have needed greater exposure to combining these skills into entire translation tasks (Hurtado Albir, 2017; Kelly, 2005).

This observation is in agreement with models of translation competence, which view translation as a multidimensional concept that goes beyond simple grammatical knowledge. Corpus tools may underpin lexical and grammatical correctness, but effective translation also depends on discourse competence, intercultural awareness, strategic decision-making, and domain-specific knowledge (Hurtado Albir, 2017; Kelly, 2005; Pym, 2010). The relatively stable performance on the translation component does not negate the usefulness of corpus-based teaching but instead shows that corpus consultation needs to be part of the overall translation pedagogy over extended instructional durations.

5.3 Evaluation of Corpus Tools and Student Perceptions

The questionnaire results also complement the quantitative results by showing that learners held remarkably positive perceptions of corpus-assisted instruction. Participants reported increased grammatical awareness and confidence in making translation decisions, as well as an enhanced capacity to verify their linguistic choices using authentic language evidence. Sketch Engine became the most frequently used corpus resource, followed closely by the BNC, which indicates that students prefer tools that give them immediate access to concordance lines, collocational information, and grammatical pattern analysis. The quantitative data are again supported by the prevalence of colligational searches among the reported corpus activities, showing that learners actively used corpus resources to explore the grammatical structures targeted during instruction (Kilgarriff et al., 2014; McEnery & Hardie, 2012).

The qualitative results also point to the pedagogical usefulness of corpus-based learning. Students repeatedly reported a gradual transition toward more evidence-based decision-making, and with repeated visits to the corpus, their confidence and independence grew. Many participants reported that the presence of authentic examples of language use allowed them to justify their translation decisions more objectively than when relying on bilingual dictionaries or teacher explanations. The findings are consistent with previous studies suggesting that corpus consultation fosters critical language awareness and motivates learner autonomy through a guided discovery approach, as opposed to merely receiving grammatical rules (Frankenberg-Garcia, 2012; Boulton & Vyatkina, 2021; Gilquin, 2021).

Meanwhile, participants pointed out some difficulties related to corpus use. The first problems were perceived with interpreting concordance lines, choosing search parameters, and understanding corpus output, especially with the use of BootCaT. However, these difficulties were overcome over time through instructor modelling, scaffolded activities, guided practice, and repeated corpus consultation. This result supports corpus-based pedagogy theories which

suggest that beginners need to be actively supported in the use of corpus tools until they are ready to work with them on their own (Vyatkina, 2016; Canning, 2014; Timmis, 2015; Crosthwaite & Wijaya, 2023). Therefore, corpus tools cannot be regarded as substitutes for instruction, but rather as resources whose pedagogical success requires a well-thought-through pedagogical integration.

There are a number of limitations that need to be noted in the interpretation of the findings. The quasi-experimental design lacked a control group, and this restricted the extent to which causal inferences could be made. The sample of twenty-one participants at a single institution is also rather small for generalizing the findings. Moreover, the twelve-week intervention period might not have been long enough to allow learners to apply newly acquired grammatical knowledge in more challenging discourse-level translation tasks. Lastly, the study considered only corpus-based instruction and did not look at the possible interaction between corpus consultation and emerging AI-assisted translation technologies.

5.4 Limitations of the Study

While this has yielded considerable insights, a few drawbacks have to be recognized. First, the quantitative results are not immediately generalizable to the broader population of Arab EFL translation students because the sample size was limited ($N=21$), and was obtained through a convenience sampling from one institution. Second, no control group was used in the study design, hence the presence of threats to internal validity. However, it is not possible to completely separate out the effects of the corpus intervention from its maturation effects (language development throughout the semester) and from its practice testing effects (familiarity with the assessment tasks). Finally, the evaluation only focused on short texts, specifically at the sentence level. Therefore, it was not possible to quantify the retention of colligational competence over time, nor was it possible to measure the extent to which the competence can be transferred to the context of translating longer, coherent discourse texts within authentic time pressure constraints.

Future research should therefore employ larger and more diverse samples across multiple educational institutions using controlled experimental or longitudinal designs. Longer instructional interventions would help determine whether improvements in grammatical pattern recognition eventually transfer to extended translation performance. Future investigations may also compare corpus-based instruction with AI-assisted translation environments, explore the relationship between corpus literacy and translator competence, and examine how corpus consultation can be integrated into translator education and teacher-training programmes (Hasumi & Chiu, 2024; Crosthwaite & Wijaya, 2023; Qiu, 2024).

6. Conclusion and Implications

6.1 Conclusion

This research investigated the effectiveness of corpus-based instruction in developing Arab EFL university students' recognition and application of English colligational grammatical patterns in translation, using a mixed-methods convergent design with a quasi-experimental one-group pre-test–post-test design. The use of the British National Corpus (BNC), Sketch Engine, and BootCaT within a DDL framework gave learners the chance to explore authentic language patterns, confirm grammatical choices, and make translation decisions using empirical linguistic evidence (Johns, 1991; McEnery & Hardie, 2012).

The results showed that corpus-based instruction had a significant beneficial effect on learners' recognition and application of English colligational grammatical patterns, with substantial gains in grammatical awareness and sentence-level translation performance. Although no statistically significant improvement was observed for extended translation tasks, the qualitative results suggest that learners expressed greater confidence, improved grammatical awareness, and greater learner autonomy while gradually shifting from intuition-based translation toward evidence-based linguistic decision-making.

These results fit in with existing research on corpus-informed pedagogy in the fields of language and translation teaching, particularly in the Arab EFL setting where empirical evidence is relatively scarce (Boulton & Vyatkina, 2021; Shawaqfeh & Khasawneh, 2023; Mekahli, 2024). The research also shows that corpus consultation can effectively supplement established translation instruction by providing genuine linguistic evidence, which reinforces grammatical competence and promotes analytical engagement with language.

6.2 Pedagogical Implications and Recommendations

The findings hold several important pedagogical implications for translation education. To start with, corpus consultation should become part of the regular translation course rather than an optional technological supplement (Bowker, 2002; Zanettin, 2009). Second, the teaching of translation should offer explicit corpus literacy training, including concordance interpretation, grammatical pattern analysis, and evidence-based decision-making (McEnery & Hardie, 2012; Timmis, 2015). Third, a scaffolded sequence of instruction that helps learners progressively move from guided corpus exploration to independent consultation seems especially useful for novice corpus users (Boulton & Vyatkina, 2021; Vyatkina, 2016). Lastly, incorporating both general reference corpora and specialized corpora can more effectively equip students for real-world professional translation work across various disciplinary settings (Baroni et al., 2004; Bowker, 2002).

The paper also extends the application of Data-Driven Learning to translation pedagogy, illustrating how corpus consultation can support the acquisition of English colligational grammatical patterns among Arab EFL students. In doing so, it adds empirical evidence to a largely underexplored area of corpus-informed translation research and demonstrates the pedagogical value of incorporating authentic corpus resources into university translation programmes.

1. Integration of Corpus Literacy in the Curriculum

Corpus consultation must be incorporated as a common, active tool into the main curriculum of translation study and cannot be simply regarded as a technological novelty, as Bowker (2002) and Zanettin (2009) have pointed out. Therefore, educators need to explicitly teach students about the concepts of corpus literacy, such as how to make queries, how to interpret the information in the lines of the concordances, how to analyze the grammatical patterns, and how to use information from the concordances to solve translation problems (McEnery & Hardie, 2012; Timmis, 2015). In order to avoid overwhelming beginners, teachers have to develop a sequence of tasks and exercises that will gradually increase in complexity: from highly structured and teacher-led, to more and more independent explorations of the corpus (Vyatkina, 2016; Boulton & Vyatkina, 2021). Moreover, the use of large general reference corpora (such as the BNC) and of specialized, domain-specific corpora will ensure that students have access to the kinds of texts that they will encounter in their future professional lives (Baroni et al., 2004; Bowker, 2002).

2. Implementation of Three-Step DDL Worksheets

It is advisable for instructors to not only correct the mistakes or mark the translations as "incorrect", but to guide the students to do the "inductive self-correction" with the active "Three-Step DDL Worksheets":

Step 1 (Identify): The instructor identifies an error that is commonly made by students in the target part of the colligation (e.g., depending of).

Step 2 (Query): The student is asked to search the corpus of the target language (e.g. BNC or Sketch Engine) for the keyword and examine his/her search results to gain insight into what it naturally collocates and the way it is used grammatically.

Step 3 (Correct & Formulate): The student writes the corrected pattern, copies three real-world concordance examples of the pattern used in context, and explicitly states the grammatical rule.

3. Development of a Digital "Corpus-to-Translation Revision Log"

Passive, traditional vocabulary notebooks should be replaced by students' "Corpus-to-Translation Revision Log". This log needs to record systematically:

The first word-for-word or grammatical mistake.

The linguistic query formulated to solve the language problem (e.g., the search query formulated for resolving the prepositional complementation issue is [prevent] [prep]).

Sample of the raw, retrieved concordance lines (with correct usage).

Final corrected translation by the student (verification).

Every week, the academic routine becomes more routine when integrating this log into it, and it becomes a natural part of the translator's work routine to verify the accuracy of the corpus, just as it is with professional translators.

6.3 Directions for Future Research

This study therefore provides empirical support for the applicability of DDL in the context of being used in Arab EFL translation programs, although further studies are required to build on these findings:

Methodological Rigor: Larger and more diverse groups of students from various academic institutions in different countries should be used in future research to confirm the pedagogical value of DDL with respect to conventional translation pedagogy with the help of rigorous randomized controlled trials (RCTs).

Longitudinal Tracking: In future designs, longitudinal timelines should be implemented to track long term cognitive retention of the competence of colligation and to see if long exposure to corpus tools improves the macro level, discourse-length translation performance.

Technological Convergence: Educational interface between corpus consultancy and the translation environment of generative AI such as LLMs or machine translation models needs to be investigated in future. The use of corpus literacy for critically checking, editing, and authenticating AI-generated translations should also be tested, and educating future translators on such an AI-enabled professional environment should also be a priority in research (Crosthwaite & Wijaya, 2023; Hasumi & Chiu, 2024; Qiu, 2024).

7. References

- Bahns, J. (1993). Lexical collocations: A contrastive view. *ELT Journal*, 47(1), 56–63. <https://doi.org/10.1093/elt/47.1.56>
- Baroni, M., Bernardini, S., Ferraresi, A., & Zanchetta, E. (2004). The BootCaT toolkit: Bootstrapping corpora and terms from the web. In M. T. Lino, M. F. Xavier, F. Ferreira, R. Costa, & R. Silva (Eds.), *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC 2004)* (pp. 1313–1316). European Language Resources Association.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Longman.
- Boonraksa, T., & Naisena, S. (2022). Corpus-based instruction and EFL learners' grammatical development. *Journal of Language Teaching and Research*, 13(4), 745–756. <https://doi.org/10.17507/jltr.1304.07>
- Boulton, A. (2010). Data-driven learning: Taking the computer out of the equation. *Language Learning*, 60(3), 534–572. <https://doi.org/10.1111/j.1467-9922.2010.00566.x>
- Boulton, A., & 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://hdl.handle.net/10125/73447>
- Bowker, L. (2002). *Computer-aided translation technology: A practical introduction*. University of Ottawa Press.

- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Brown, H. D. (2006). *Principles of language learning and teaching* (5th ed.). Pearson Education.
- Burnard, L. (2007). *Reference guide for the British National Corpus (XML edition)*. Oxford University Computing Services.
- Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language*, 82(4), 711–733. <https://doi.org/10.1353/lan.2006.0186>
- Canning, J. (2014). *Teaching language with corpora*. Palgrave Macmillan.
- Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). Sage.
- Crosthwaite, P., & Wijaya, T. (2023). Corpus-informed pedagogy and language learning in the age of AI. *Language Learning & Technology*, 27(2), 1–18. <https://doi.org/10.1017/S026144482300021X>
- El-Dakhs, D. A. S. (2015). The integration of corpus linguistics into translator training. *International Journal of English Language Teaching*, 3(5), 1–18.
- Fang, F., Zhang, T., & Li, X. (2021). Corpus-based learning and its impact on lexical quality in translation. *International Journal of Applied Linguistics*, 31(2), 202–216. <https://doi.org/10.1111/ijal.12331>
- Farghal, M., & Shunnaq, A. (1999). *Translation with reference to English and Arabic: A practical guide*. Irbid National University.
- Firth, J. R. (1957). *Papers in linguistics, 1934–1951*. Oxford University Press.
- Flowerdew, L. (2015). Corpus-based research and pedagogy in EAP: From lexis to genre. *Language Teaching*, 48(1), 99–116. <https://doi.org/10.1017/S0261444813000037>
- Frankenberg-Garcia, A. (2012). Learners' use of corpora for preposition choice. *Language Awareness*, 21(1–2), 27–43. <https://doi.org/10.1080/09658416.2011.639885>
- Gilquin, G. (2021). Using corpora to foster L2 construction learning: A data-driven learning experiment. *International Journal of Applied Linguistics*, 31(2), 229–247. <https://doi.org/10.1111/ijal.12317>
- Hasumi, T., & Chiu, M. S. (2024). Technology-enhanced language learning in English language education: Performance analysis, core publications, and emerging trends. *Cogent Education*, 11(1), Article 2346044. <https://doi.org/10.1080/2331186X.2024.2346044>
- Hoey, M. (2005). *Lexical priming: A new theory of words and language*. Routledge.
- Howarth, P. (1996). *Phraseology in English academic writing: Some implications for language learning and dictionary making*. Niemeyer. <https://doi.org/10.1515/9783110937923>.
- Hunston, S., & Francis, G. (2000). *Pattern grammar: A corpus-driven approach to the lexical grammar of English*. John Benjamins.
- Hurtado Albir, A. (2017). Translation and translation competence. In A. Hurtado Albir (Ed.), *Researching translation competence by PACTE Group* (Benjamins Translation Library, Vol. 127, pp. 3–34). John Benjamins. <https://doi.org/10.1075/btl.127.01hur>

- Johns, T. (1991). Should you be persuaded: Two samples of data-driven learning materials. In T. Johns & P. King (Eds.), *Classroom concordancing* (ELR Journal, Vol. 4, pp. 1–16). Centre for English Language Studies, University of Birmingham.
- Katan, D. (2004). *Translating cultures: An introduction for translators, interpreters, and mediators*. Routledge.
- Kelly, D. (2005). *A handbook for translator trainers: A guide to reflective practice*. St. Jerome Publishing.
- Keshavarz, M. H., & Salimi, H. (2007). Collocational competence and translation performance. *Journal of Applied Linguistics*, 2(1), 59–76.
- Kilgariff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). The Sketch Engine: Ten years on. *Lexicography*, 1(1), 7–36. <https://doi.org/10.1007/s40607-014-0009-9>
- Lado, R. (1957). *Linguistics across cultures: Applied linguistics for language teachers*. University of Michigan Press.
- Mahmoud, A. (2005). Collocation errors made by Arab learners of English. *Asian EFL Journal*, 5(2), 117–126.
- McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.
- McEnery, T., & Wilson, A. (2001). *Corpus linguistics: An introduction*. Edinburgh University Press.
- McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. Routledge.
- Mekahli, S. (2024). Translation strategies of Arabic and English collocations: A case study of Algerian EFL learners. *Dirasat: Human and Social Sciences*, 51(2), 321–332. <https://doi.org/10.35516/hum.v51i2.3718>
- Metwally, A. A., & Hamad, M. (2023). The impact of collocational competence on translation quality: A literature review. *AWEJ for Translation & Literary Studies*, 7(1), 88–102. <https://doi.org/10.2139/ssrn.4386585>
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge University Press.
- Nation, I. S. P. (2009). *Teaching ESL/EFL reading and writing*. Routledge.
- Nesselhauf, N. (2005). *Collocations in a learner corpus*. John Benjamins. <https://doi.org/10.1075/scl.14>
- Newmark, P. (2001). *A textbook of translation*. Shanghai Foreign Language Education Press.
- O'Keeffe, A., & McCarthy, M. (2010). From corpus to classroom: Language use and language teaching. In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 563–571). Routledge.
- Odlin, T. (1989). *Language transfer: Cross-linguistic influence in language learning*. Cambridge University Press.
- Pym, A. (2010). *Exploring translation theories*. Routledge.
- Qiu, X. (2024). Exploring the effect of corpus-based writing instruction on learner-corpus interaction in L2 revision: A study of Chinese EFL disciplinary writers. *TESOL Quarterly*, 58(3), 1108–1137. <https://doi.org/10.1002/tesq.3241>

- Shawaqfeh, A. T., & Khasawneh, M. A. S. (2023). Incorporating corpus linguistics tools in the training and professional development of lecturers in translation studies. *Studies in Media and Communication*, 11(7), 263–270. <https://doi.org/10.11114/smc.v11i7.6412>
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.
- Sinclair, J. (Ed.). (2004). *How to use corpora in language teaching*. John Benjamins.
- Sun, W., & Park, E. (2023). EFL learners' collocation acquisition and learning in corpus-based instruction: A systematic review. *Sustainability*, 15(18), Article 13242. <https://doi.org/10.3390/su151813242>
- Timmis, I. (2015). *Corpus linguistics for ELT: Research and practice*. Routledge.
- Tsai, M. H. (2020). The effects of explicit instruction on L2 learners' acquisition of verb–noun collocations. *Language Teaching Research*, 24(2), 138–162. <https://doi.org/10.1177/1362168818795188>
- Uçar, S., & Yükselir, C. (2015). The effect of corpus-based activities on verb–noun collocations in EFL classes. *International Journal of Educational Research*, 71, 22–31. <https://doi.org/10.1016/j.ijer.2015.01.005>
- Venuti, L. (1995). *The translator's invisibility: A history of translation*. Routledge.
- Vyatkina, N. (2016). Data-driven learning of collocations: Learner performance, proficiency, and perceptions. *Language Learning & Technology*, 20(3), 159–179. <https://hdl.handle.net/10125/44485>
- Yılmaz, E., & Soruç, A. (2015). The use of concordance for teaching vocabulary: A data-driven learning approach. *Journal of Educational Sciences*, 15(3), 156–173.
- Yoon, H., & Hirvela, A. (2004). ESL student attitudes toward corpus use in L2 writing. *Journal of Second Language Writing*, 13(4), 257–283. <https://doi.org/10.1016/j.jslw.2004.06.002>
- Zaki, M. (2021). Corpus-based language teaching and learning: Applications and implications. *International Journal of Applied Linguistics*, 31(2), 169–172. <https://doi.org/10.1111/ijal.12316>
- Zanettin, F. (2009). Corpus-based translation activities for language learners. *The Interpreter and Translator Trainer*, 3(2), 217–235. <https://doi.org/10.1080/1750399X.2009.10798789>

8. **Appendixes**

Appendix A – Student Questionnaire

Student Perceptions of Using Corpus-Based Tools in Learning English Colligational Patterns for Translation

Evaluating the Use of BNC, Sketch Engine, and BootCaT in Enhancing Translation Skills

Introduction:

You are invited to take part in a research study about the use of corpus-based tools in learning Arabic-English translation.

Purpose of the Study:

This questionnaire aims to understand your experience, challenges, and opinions about using tools like the British National Corpus (BNC), Sketch Engine, and BootCaT. The study focuses on how these tools help with learning colligational patterns and improving translation skills.

Participation Information:

Your answers will be used for research only.

- All responses are anonymous and confidential.
- There are no right or wrong answers.
- Participation is voluntary. You may skip any question or stop at any time.

Instructions:

- Answer all questions honestly based on your experience.
- For rating questions, choose the option that best reflects your view.
- Use the open-ended sections to share more detail if you wish.

Your participation is appreciated and will help improve translation training methods.

The questionnaire consists of 30 questions, divided into sections. It will take 10–15 minutes to complete.

Check the appropriate box or write your response clearly in the space provided.

Thank you for your time and cooperation.

Section 1: Background Information

1. What is your age?

- ☐ 18
- ☐ 19
- ☐ 20

2. What is your gender?

- ☐ Male
- ☐ Female
- ☐ Prefer not to say

3. How would you rate your general English proficiency?

- ☐ Beginner
- ☐ Elementary
- ☐ Intermediate
- ☐ Upper - intermediate or higher

4. What type of translation course are you currently taking?

- ☐ General
- ☐ Specialized (Legal-Un, etc.)
- ☐ other

5. Have you ever taken any formal training or workshop on corpus use before this course?

- ☐ Yes
- ☐ No

6- Before this course, had you ever used the British National Corpus (BNC)?

- ☐ Yes
- ☐ No

7- Have you ever used Sketch Engine?

- ☐ Yes
- ☐ No

8- Have you ever used BootCaT?

- ☐ Yes
- ☐ No

9- Have you ever used any language corpora before this training?

- ☐ Yes
- ☐ No

10- Do you have a clear understanding of what a language corpus is after finishing this training?

- ☐ Yes
- ☐ No

Section 2: Training and usage

11- Which tool did you use most frequently?

- ☐ BNC
- ☐ Sketch engine
- ☐ BootCaT

12- Which of these functions did you try? (check all that apply)

- ☐ Searching for collocations
- ☐ Checking grammar patterns
- ☐ Building a web corpus
- ☐ Comparing synonym usage

13- How often did you use the corpus tools during the training?

- ☐ Rarely
- ☐ Sometimes
- ☐ Often
- ☐ Every session

14- Did you face any difficulties while using the tools?

- ☐ Yes
- ☐ No

15- Were the training instructions clear and easy to follow?

- ☐ Yes
- ☐ No

16- The tools helped me learn how to use prepositions more accurately

- ☐ Strongly disagree
- ☐ Disagree
- ☐ Neutral
- ☐ Agree
- ☐ Strongly agree

17- The tools improved my understanding of sentence structures

- ☐ Strongly disagree
- ☐ Disagree
- ☐ Neutral
- ☐ Agree
- ☐ Strongly agree

18- I used corpus data to revise or change a translation decision

- ☐ Strongly disagree
- ☐ Disagree
- ☐ Neutral
- ☐ Agree
- ☐ Strongly agree

19- I became more aware of how colligations affect translation

- ☐ Strongly disagree
- ☐ Disagree
- ☐ Neutral
- ☐ Agree
- ☐ Strongly agree

20- The corpus helped me avoid literal translation errors.

- ☐ Strongly disagree
- ☐ Disagree
- ☐ Neutral
- ☐ Agree
- ☐ Strongly agree

21- I now pay more attention to context when choosing colligational grammatical patterns.

- ☐ Strongly disagree
- ☐ Disagree
- ☐ Neutral
- ☐ Agree
- ☐ Strongly agree

22- Using corpus tools is more reliable than using only dictionaries.

- ☐ Strongly disagree
- ☐ Disagree
- ☐ Neutral
- ☐ Agree
- ☐ Strongly agree

23- I feel more confident in making translation choices after the training.

- ☐ Strongly disagree
- ☐ Disagree
- ☐ Neutral
- ☐ Agree
- ☐ Strongly agree

24- I believe corpus tools should be part of all translation courses.

- ☐ Strongly disagree
- ☐ Disagree
- ☐ Neutral
- ☐ Agree
- ☐ Strongly agree

25- I plan to use Sketch Engine or BNC in the future for my assignments.

- ☐ Strongly disagree
- ☐ Disagree
- ☐ Neutral
- ☐ Agree
- ☐ Strongly agree

26- The tools were useful but difficult to use.

- ☐ Strongly disagree
- ☐ Disagree
- ☐ Neutral
- ☐ Agree
- ☐ Strongly agree

27- I prefer using real language data over guessing in translation tasks

- ☐ Strongly disagree
- ☐ Disagree
- ☐ Neutral
- ☐ Agree
- ☐ Strongly agree

28- The corpus tools helped me recognize and apply correct colligational patterns in my translations.

- ☐ Strongly disagree
- ☐ Disagree
- ☐ Neutral
- ☐ Agree
- ☐ Strongly agree

Section 5: Feedback and suggestions

29- What part of the corpus training helped you the most in understanding or using colligational patterns?

.....

30- What would you improve in the corpus-based training, especially regarding colligational pattern practice, for future students?

Appendix B – Pre- and Post-Tests (Questions and sample of students' answers)

The pre- and post-tests were distributed to the students via Google Forms. Students submitted their responses through the same platform.

First: pre-test questions:

**Pre-Test on Colligational Grammatical Patterns
For Translation Students – AOU Egypt**

Purpose

This test assesses your baseline proficiency in using English **colligational grammatical patterns**. Your responses will help evaluate your understanding and application of these structures in translation.

Instructions

- Time allowed: 75 minutes
- Answer all questions
- Read each item carefully
- Use correct **colligational grammatical patterns** in all responses

Section 1: Colligational Grammatical Pattern Identification

Choose the correct preposition to complete each sentence. Mark only one answer.

1. He is afraid ____ making mistakes.
 - a) from
 - b) of
 - c) in
 - d) to
2. She is good ____ solving problems.
 - a) at
 - b) on
 - c) in
 - d) to
3. They succeeded ____ passing the exam.
 - a) to

- b) on
 - c) in
 - d) at
4. She is interested ____ art.
- a) with
 - b) in
 - c) to
 - d) on
5. He apologized ____ being late.
- a) for
 - b) to
 - c) with
 - d) on
6. We depend ____ their support.
- a) in
 - b) on
 - c) with
 - d) to
7. He insisted ____ seeing her.
- a) to
 - b) on
 - c) at
 - d) with
8. She is capable ____ handling the project.
- a) of
 - b) to

- c) on
 - d) in
9. They were accused ____ stealing the documents.
- a) of
 - b) with
 - c) on
 - d) at
10. He was responsible ____ organizing the event.
- a) on
 - b) in
 - c) for
 - d) to
11. I object ____ his opinion on the issue.
- a) to
 - b) on
 - c) at
 - d) in
12. She insisted ____ coming with us.
- a) at
 - b) with
 - c) on
 - d) by
13. We are committed ____ improving our services.
- a) in
 - b) to
 - c) for

- d) at

14. He congratulated me ____ my success.

- a) of
- b) for
- c) with
- d) on

15. They accused him ____ cheating in the exam.

- a) for
- b) with
- c) of
- d) on

16. He was tired ____ working late.

- a) on
- b) of
- c) to
- d) for

17. She apologized ____ the inconvenience.

- a) on
- b) to
- c) for
- d) with

18. He succeeded ____ convincing the client.

- a) in
- b) to
- c) for
- d) on

19. They insisted ____ attending the meeting.

- a) on
- b) in
- c) to
- d) with

20. We relied ____ accurate information.

- a) with
- b) on
- c) for
- d) at

21. The manager is known ____ his fairness.

- a) by
- b) of
- c) for
- d) at

22. She is responsible ____ the budget.

- a) on
- b) for
- c) in
- d) to

23. They objected ____ the policy.

- a) at
- b) on
- c) to
- d) with

24. He is committed ____ his work.

- a) to
- b) on
- c) in
- d) at

25. He is interested ____ linguistics.

- a) in
- b) to
- c) for
- d) on

Section 2: Colligational Grammatical Pattern Application

Fill in the blank with the correct preposition to complete the colligational grammatical pattern.

- 26 – He succeeded ____ solving the problem
- 27 – She is afraid ____ making mistakes in translation
- 28 – They insisted ____ finishing the project
- 29 – The professor is known ____ his work in translation theory
- 30 – He apologized ____ the delay
- 31 – She is capable ____ leading the team
- 32 – We depend ____ accurate data
- 33 – He objected ____ the changes
- 34 – I am interested ____ reading about corpus linguistics
- 35 – She is responsible ____ the final draft
- 36 – They succeeded ____ reaching a conclusion
- 37 – She is good ____ managing her time
- 38 – We rely ____ up-to-date references

- 39 – He is committed ___ the goals of the team
- 40 – He insisted ___ helping her
- 41 – He is capable ___ handling complex tasks
- 42 – We object ___ the new policy
- 43 – She was tired ___ waiting for the bus
- 44 – I am committed ___ improving my skills
- 45 – They insisted ___ discussing the issue further
- 46 – She is interested ___ learning new languages
- 47 – We depend ___ the accuracy of the data
- 48 – He apologized ___ the misunderstanding
- 49 – She is good ___ negotiating deals
- 50 – We rely ___ their expertise
- 51 – He succeeded ___ achieving the desired outcome
- 52 – She is responsible ___ the team’s progress
- 53 – He was known ___ his innovative approach
- 54 – They objected ___ the decision made by the committee
- 55 – She is committed ___ delivering high-quality work
- 56 – We insisted ___ staying late to finish the task
- 57 – He apologized ___ his absence from the meeting
- 58 – She succeeded ___ completing the project on time
- 59 – I am interested ___ exploring different methodologies
- 60 – They were known ___ their contributions to the field

Section 3: Translation Tasks

Translate each Arabic sentence into English. Use correct colligational grammatical patterns.

- 61 - يَعْتَمِدُ النَّجَاحُ عَلَى التَّخَطُّيطِ الْجَيِّدِ -
- 62 - اعْتَذَرَ عَنِ التَّأخِيرِ فِي السَّلَامِ -
- 63 - هُوَ مَعْرُوفٌ بِإِنْجَازِهِ الْمُهَيَّي -
- 64 - هِيَ مُهْتَمَّةٌ بِالتَّنْزِيلِ عَلَى التَّرْجَمَةِ -
- 65 - هُوَ مَسْئُولٌ عَنْ إِعْزَازِ النَّقَارِ -
- 66 - نَحْنُ نَعْتَمِدُ عَلَى مَوَادِّ مَوْثُوقَةٍ -
- 67 - أَصَرَ عَلَى الْمَشَارَكَةِ فِي الْمُؤْتَمَرِ -
- 68 - هِيَ جَيِّدَةٌ فِي تَنْطِيمِ الْوَقْتِ -
- 69 - أَنَا مُلْتَزِمٌ بِتَطْوِيرِ مَهَارَاتِي -
- 70 - هُوَ يُخْتَصُّ فِي التَّرْجَمَةِ الْقَانُونِيَّةِ -
- 71 - هُوَ مُهْتَمٌّ بِتَطْوِيرِ نَظَرِيَّاتِ التَّرْجَمَةِ -
- 72 - أَنَا مَسْئُولٌ عَنْ تَجْهِيْزِ الْوُثَائِقِ -
- 73 - نَحْنُ نَعْتَمِدُ عَلَى التَّحْلِيلِ اللَّغَوِيِّ -
- 74 - هِيَ مُلْتَزِمَةٌ بِمُرَاجَعَةِ الْمَشَارِيحِ -
- 75 - أَصَرَ عَلَى تَجْرِبَةِ طَرِيقَةٍ جَدِيدَةٍ -
- 76 - هُوَ مَعْرُوفٌ بِكِفَائَتِهِ الْعِلْمِيَّةِ -
- 77 - هِيَ مُهْتَمَّةٌ بِقِرَاءَةِ الْأَبْحَاثِ الْأَكَادِيمِيَّةِ -
- 78 - هُوَ مَسْئُولٌ عَنْ تَنْمِيقِ الْفَرْقِ الرِّيَاضِيَّةِ -
- 79 - نَحْنُ نَعْتَمِدُ عَلَى الْمُعَاجِمِ الْإِلِكْتُرُونِيَّةِ -
- 80 - أَنَا مُلْتَزِمٌ بِتَطْوِيرِ مَنَهِجِيَّيِ الْبَحْثِيَّةِ -

Second: Post-test questions :

POST-TEST

Proficiency in Colligational Grammatical Patterns

For Translation Students – AOU Egypt

Purpose

This test measures your proficiency in using English **colligational grammatical patterns** after corpus-based training.

Instructions

- Time allowed: 75 minutes
 - Answer all questions
 - Read each item carefully
 - Use correct **colligational grammatical patterns** in all responses
-

Section 1: Colligational Grammatical Pattern Identification

Choose the correct preposition to complete each sentence. Mark only one answer.

1. They agreed ____ the new proposal.
 - a) with
 - b) on
 - c) about
 - d) for
2. She insisted ____ paying the full amount.
 - a) on
 - b) in
 - c) at
 - d) to

3. He is capable ____ handling complex tasks.
- a) of
 - b) in
 - c) for
 - d) to
4. We are interested ____ expanding our vocabulary.
- a) on
 - b) in
 - c) with
 - d) at
5. The team was responsible ____ organizing the event.
- a) for
 - b) of
 - c) in
 - d) on
6. I am tired ____ repeating the same mistakes.
- a) for
 - b) of
 - c) with
 - d) at
7. She apologized ____ the inconvenience caused.
- a) to
 - b) for
 - c) on
 - d) with
8. He succeeded ____ convincing the client.

- a) in
- b) on
- c) at
- d) of

9. They objected ____ the unfair decision.

- a) to
- b) on
- c) at
- d) for

10. We rely ____ accurate information for our research.

- a) on
- b) at
- c) in
- d) for

11. He is known ____ his expertise in translation studies.

- a) for
- b) of
- c) by
- d) at

12. She is committed ____ improving her language skills.

- a) to
- b) in
- c) on
- d) at

13. The professor congratulated him ____ winning the award.

- a) on

- b) for
- c) at
- d) in

14. I object ____ changing the course syllabus.

- a) at
- b) to
- c) on
- d) with

15. They insisted ____ providing additional training.

- a) on
- b) in
- c) at
- d) for

16. He is responsible ____ submitting the final draft.

- a) for
- b) to
- c) in
- d) on

17. She depends ____ reliable data sources.

- a) in
- b) on
- c) at
- d) for

18. The students succeeded ____ completing the project on time.

- a) at
- b) on

- c) in
- d) for

19. We apologized ____ the delay in responses.

- a) for
- b) to
- c) on
- d) with

20. He insisted ____ attending the workshop.

- a) on
- b) in
- c) at
- d) for

21. She is interested ____ corpus linguistics research.

- a) at
- b) in
- c) on
- d) for

22. They were accused ____ plagiarism in the paper.

- a) of
- b) for
- c) by
- d) on

23. I am committed ____ applying corpus methods in teaching.

- a) to
- b) on
- c) in

- d) at

24. He objected ____ the proposed amendments.

- a) to
- b) on
- c) at
- d) for

25. The manager congratulated the team ____ successful completion.

- a) on
- b) for
- c) of
- d) at

Section 2: Colligational Grammatical Pattern Application

Fill in the blank with the correct preposition to complete each colligational grammatical pattern.

26. She succeeded ____ improving her translation accuracy.

27. He apologized ____ missing the meeting.

28. They insisted ____ receiving feedback regularly.

29. We depend ____ authentic sources for reference.

30. He is responsible ____ monitoring student progress.

31. She is interested ____ exploring new linguistic tools.

32. The group objected ____ the changes in deadlines.

33. I am committed ____ completing my thesis on time.

34. The professor congratulated the students ____ their research achievements.

35. He insisted ____ attending the conference next month.

36. She apologized ____ the technical difficulties experienced.

37. They succeeded ____ collaborating effectively on the project.

38. We rely ___ up-to-date software for corpus analysis.
39. The team objected ___ the lack of clear instructions.
40. He is capable ___ managing complex translation projects.
41. I depend ___ my supervisor's guidance.
42. She is committed ___ adopting new teaching methods.
43. They apologized ___ any inconvenience caused by the delay.
44. He objected ___ the proposed grading scheme.
45. He agreed ___ the teacher's feedback.
46. The students are interested ___ applying corpus tools in translation.
47. She is responsible ___ organizing the final presentation.
48. They insisted ___ revising the draft thoroughly.
49. I am committed ___ enhancing my language proficiency.
50. He apologized ___ the confusion in the instructions.
51. The group succeeded ___ meeting all project deadlines.
52. She depends ___ accurate data for her analysis.
53. We objected ___ the cancellation of the workshop.
54. He is known ___ his expertise in corpus linguistics.
55. The manager congratulated the staff ___ excellent performance.
56. She insisted ___ reviewing all sources carefully.
57. I apologize ___ any misunderstanding.
58. They succeeded ___ producing a high-quality report.
59. He is interested ___ further research in translation studies.
60. We rely ___ feedback to improve our teaching.

Section 3: Translation Tasks

Translate each Arabic sentence into English. Use correct colligational grammatical patterns.

61. يَعْتَمِدُ النَّجَاحُ عَلَى الدِّرَاسَةِ الْمُتَمَكِّنَةِ.
62. أَصْرٌ عَلَى الْخُصُورِ إِلَى الْبَقَاءِ.
63. هِيَ مَسْئُولَةٌ عَنْ تَحْضِيرِ الْمُخَاضِرَاتِ.
64. نَحْنُ نَعْتَمِدُ عَلَى الْمَصَادِرِ الْمُوثُوقَةِ.
65. أَنَا مُلْتَزِمٌ بِتَحْسِينِ مَهَارَاتِي.
66. هُوَ مَعْرُوفٌ بِمَهَارَتِهِ فِي التَّرْجُمَةِ.
67. هِيَ مُهْتَمَّةٌ بِدِرَاسَةِ اللُّغَةِ الْإِنْجِلِيزِيَّةِ.
68. نَحْنُ نَعْتَمِدُ عَلَى تَحْلِيلِ النُّصُوصِ.
69. أَصْرٌ عَلَى سَلِيمِ الْمَشْرُوعِ فِي وَفَيْهِ.
70. هِيَ جَيِّدَةٌ فِي تَنْظِيمِ الْفَرِيقِ.
71. أَنَا أَعْتَذِرُ عَنْ تَأْخِيرِ الْمُسْتَقْدَاتِ.
72. هُوَ مَسْئُولٌ عَنْ تَقْيِيمِ الْأَدَاءِ.
73. نَحْنُ مُلْتَزِمُونَ بِتَطْوِيرِ مَهَارَاتِنَا.
74. هِيَ تَعْتَمِدُ عَلَى مَوَادِّ مُخْتَارَةٍ.
75. أَنَا أَعْتَذِرُ عَنِ الْخَلَلِ الْقَلْبِيِّ.
76. هُوَ مَعْرُوفٌ بِجَيِّزَتِهِ فِي التَّحْلِيلِ.
77. نَحْنُ نَعْتَمِدُ عَلَى أُدِلَّةٍ مُوثُوقَةٍ.
78. هِيَ مُلْتَزِمَةٌ بِمَرَاجَعَةِ الْوَثَائِقِ.
79. أَصْرٌ عَلَى التَّحْقِيقِ فِي الْمَشْكَلَةِ.
80. هُوَ يَمْتَلِكُ مَهَارَةً فِي إِدَارَةِ الْمَشَارِيعِ.

Table 12. Section 3: Translation Tasks - Grading Rubric

Criteria	Excellent (4 Points)	Good (3 Points)	Fair (2 Points)	Needs Improvement (1 Point)	Unsatisfactory (0 Points)
Use of Colligational Patterns	Correct use of prepositions in all sentences.	Mostly correct use of prepositions, with a few errors.	Some correct use of prepositions, but many errors are present.	Incorrect use of prepositions throughout most sentences.	No awareness of required colligational patterns.
Clarity and Cohesion	Translations are clear, coherent, and very readable.	Translations are mostly clear and coherent.	Some clarity issues, and translations may be hard to follow at times.	Translations lack clarity and cohesion, making them difficult to understand.	Translations are unclear and incoherent.
Grammar and Syntax	No grammatical or syntactical errors.	A few grammatical or	Several errors, but they do not	Numerous errors significantly affect readability	Frequent errors make the

		syntactical errors are present.	interfere with the meaning.	and understanding.	sentences unreadable.
--	--	------------------------------------	--------------------------------	-----------------------	--------------------------

Plan outline

Step-by-Step Implementation Plan for the Practical Methodology

This plan outlines the steps you need to follow when applying your proposal in the classroom, from introducing corpus-based tools to assessing student progress. The **12-week timeline** aligns with your research design and ensures structured training for students.

Week 1-2: Introduction to the Study and Training on Corpus Tools

Step 1: Inform Students About the Study (Day 1)

- Explain the **purpose of the study**:
 - Why colligational structures are essential in translation.
 - How corpus-based tools can help them recognize and apply correct colligational patterns.
- Outline their **role and expectations**:
 - They will participate in structured training sessions.
 - They will complete translation tasks using corpus tools.
 - Their progress will be assessed through pre- and post-tests.

Step 2: Conduct the Pre-Test (Day 2-3)

- Administer the **pre-test** to assess baseline proficiency.
- Tests will include:
 - Colligational identification (MCQs).
 - Sentence completion.
 - Arabic-English translation tasks.

Step 3: Introduce Corpus-Based Tools (Day 4-6)

- **Demonstrate the British National Corpus (BNC), Sketch Engine, and BootCaT.**

- Provide **step-by-step guidance** on:
 - How to search for colligational structures.
 - How to interpret concordance lines.
 - How to compare word patterns across different contexts.
- Assign **small exploratory tasks** (e.g., "Use Sketch Engine to find collocations for the word 'depend'").

Week 3-6: Hands-On Training and Guided Activities

Step 4: Supervised Practice on Colligations (Week 3-4)

- Conduct **guided exercises** where students:
 - Use BNC and Sketch Engine to analyze colligations.
 - Compare corpus data with traditional grammar rules.
 - Discuss patterns they observe.
- Assign **translation exercises** incorporating corpus-based findings.
- Provide **weekly feedback** on students' performance.

Step 5: Independent Use of Corpus Tools (Week 5-6)

- Assign **individual tasks**, such as:
 - Analyzing real-world examples of preposition usage.
 - Translating sentences using corpus data as justification.
- Monitor **common errors** and provide corrections.
- Discuss **challenges students face** in using corpus tools.

Week 7-10: Real-World Application of Corpus-Based Learning

Step 6: Advanced Translation Tasks (Week 7-8)

- Assign **longer, more complex texts** requiring colligational accuracy.

- Have students:
 - Use corpus tools to verify their translation choices.
 - Provide explanations for their word choices using corpus evidence.

Step 7: Peer Review and Discussion (Week 9-10)

- Pair students and have them **review each other's translations**.
- Encourage **justifications using corpus-based findings**.
- Organize **group discussions** on translation challenges.

Week 11-12: Assessment and Feedback

Step 8: Conduct the Post-Test (Week 11)

- Administer the **same test format** as the pre-test.
- Include a **corpus-based task** where students must:
 - Use a corpus tool to justify a translation.

Step 9: Collect Student Feedback (Week 12)

- Distribute the **questionnaire** to assess student perceptions.
- Conduct **interviews** with a sample of students to gather qualitative insights.

Step 10: Analyze Data and Report Findings

- Compare **pre-test and post-test results**.
- Identify **patterns in student performance and errors**.
- Analyze **questionnaire and interview responses** to evaluate student engagement and difficulties.

Appendix D– Training Materials and Follow-up exercises with students' sample answers Screenshot 1: BNC's interface

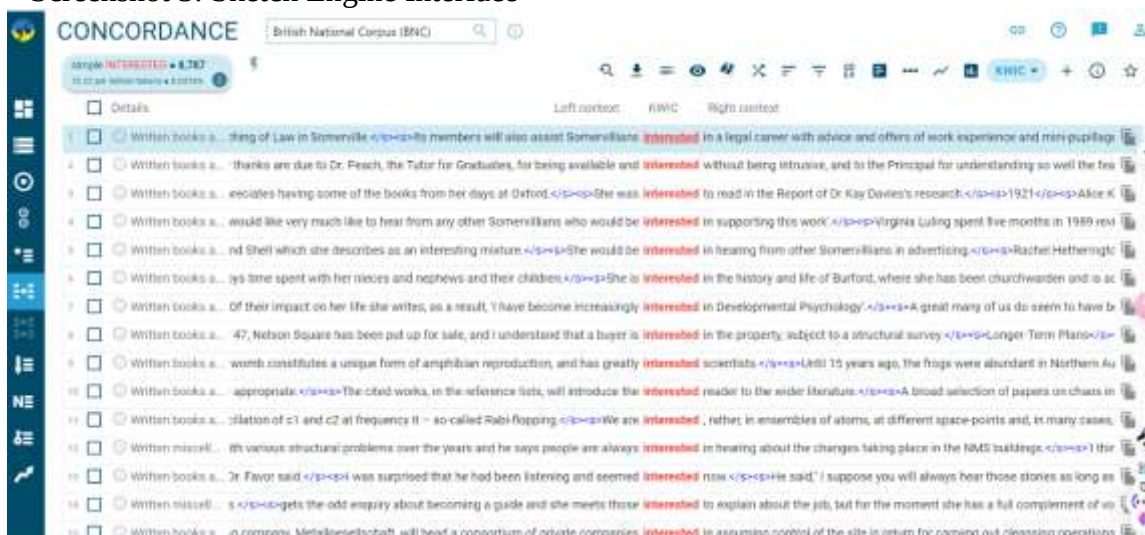


Screenshot 2: how the corpus can be searched.



Sketch Engine will be used to analyze word combinations like verb–preposition, noun-preposition, and adjective-preposition combinations, with the Word Sketch feature assisting in identifying frequent and contextually appropriate colligational grammatical patterns.

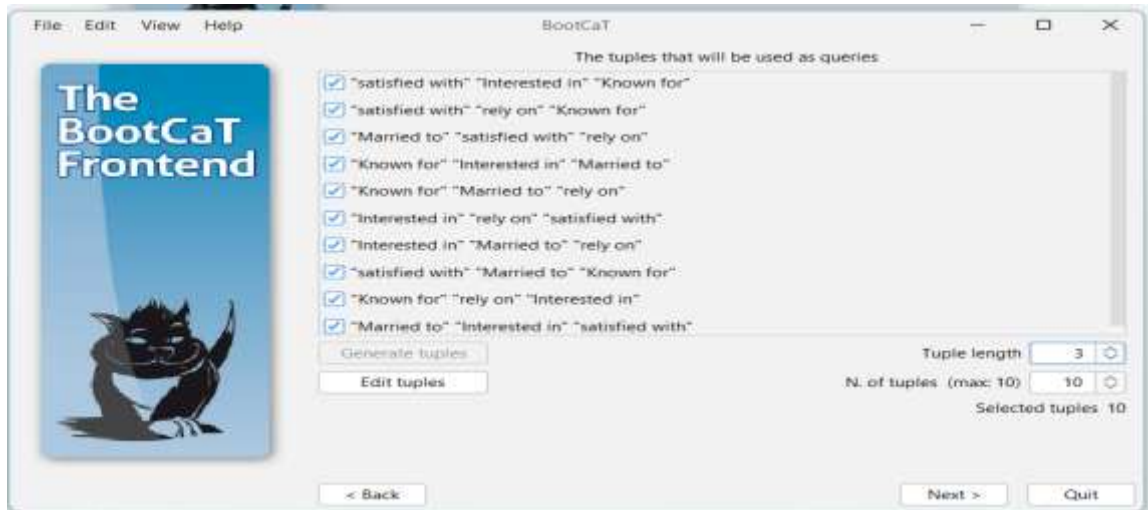
Screenshot 3: Sketch Engine Interface



Screenshot 4: Example of list of collocations.

Specific data from the web is then analyzed in Sketch Engine.

Screenshot 5 : BootCaT interface



Screenshots showing how to input keywords and generate a custom corpus using BootCaT.

Appendix H. – Common Prepositional Error Types in Student Responses

Mistake Type	Example	Explanation	Correction
Wrong Preposition Choice	He is afraid from making mistakes.	Arabic uses "من" (min, from) with "afraid", but English requires "of".	He is afraid of making mistakes.
Omitting the Preposition	She is good solving problems.	"Good" requires "at" before the gerund in English.	She is good at solving problems.
Adding Unnecessary Prepositions	They succeeded to passing the exam.	"Succeed" takes "in" + gerund.	They succeeded in passing the exam.
Transferring Arabic Patterns	He apologized about being late.	Arabic may use "عن" (about), but English requires "for".	He apologized for being late.
Confusing Similar Prepositions	We depend in their support.	Correct phrase is "depend on", not "in".	We depend on their support.
Using "for" Instead of "to"	They objected for the policy.	The correct preposition after "object" is "to".	They objected to the policy.
Using "on" Instead of "in" or "at"	She is interested on art.	"Interested in" is the correct form.	She is interested in art.
Omitting Prepositions with Verbs	He insisted seeing her.	"Insist on" is needed before a gerund.	He insisted on seeing her.
Incorrect Preposition with "responsible"	He was responsible on organizing the event.	"Responsible for" is the correct form.	He was responsible for organizing the event.
Mixing Up "of" and "for"	She is capable for handling the project.	"Capable of" is the correct structure.	She is capable of handling the project.

Source: Pre-test analysis of Section 1 responses.

Appendix I: Transcriptions of Audio Feedback (WhatsApp Voice Messages)

1. Appendix I presents anonymized transcripts of students' WhatsApp voice messages, which were collected as part of the qualitative data set. The excerpts are provided in their original language with English translations where applicable. All participants are identified by codes (S1–S10) to ensure anonymity.

2. Transcripts (sample format):

Participant Code	Original Message (Arabic/English)	English Translation (if applicable)
S1	ساعدني لاحظ Sketch Engine بصراحة، استخدام الأخطاء في الترجمات التي كنت يعملها.	Honestly, using Sketch Engine helped me notice the mistakes I used to make in my translations.
S2	The BNC examples showed me how native speakers use prepositions, which improved my confidence.	—
S3	كان صعب في الأول، لكن لما فهمت، بقيت BootCaT. اعرف اجيب امثلة للترجمة التخصصية بسهولة.	BootCaT was difficult at first, but once I understood it, I could easily find examples for specialized translation.

Note: Only representative excerpts are included here for brevity; full transcripts are available upon request.