

تحليل المكاسب النحوية من خلال التعلم القائم على المدونات: دراسة حالة وتحليل للأخطاء
في ترجمات عربية-إنجليزية لخريجي اللسانيات

محمود عبد الخالق يحيى البكوع
أستاذ مساعد / جامعة الموصل/ كلية التربية للبنات

البريد الإلكتروني Email : m.albagooh@uomosul.edu.iq

الكلمات المفتاحية: التعلم القائم على المدونات، النحو، الترجمة عربي-إنجليزي، تحليل الأخطاء.

كيفية اقتباس البحث

البكوع ، محمود عبد الخالق يحيى، تحليل المكاسب النحوية من خلال التعلم القائم على
المدونات: دراسة حالة وتحليل للأخطاء في ترجمات عربية-إنجليزية لخريجي اللسانيات، مجلة
مركز بابل للدراسات الانسانية، تموز 2026، المجلد: 16، العدد: 7.

هذا البحث من نوع الوصول المفتوح مرخص بموجب رخصة المشاع الإبداعي لحقوق التأليف
والنشر (Creative Commons Attribution) تتيح فقط للآخرين تحميل البحث
ومشاركته مع الآخرين بشرط نسب العمل الأصلي للمؤلف، ودون القيام بأي تعديل أو
استخدامه لأغراض تجارية.

مسجلة في
ROAD

مفهرسة في
IASJ



Analyzing Syntactic Gains through Corpus-Based Learning: A Case Study and Error Analysis of Arabic-English Translations by Linguistics Graduates

Mahmood Abdulkhaliq Al-Bagoa

Assistant Professor/ University of Mosul / Education college for Women

Keywords : Corpus-Based Learning, Syntax, Arabic-English Translation, Error Analysis.

How To Cite This Article

Al-Bagoa, Mahmood Abdulkhaliq, Analyzing Syntactic Gains through Corpus-Based Learning: A Case Study and Error Analysis of Arabic-English Translations by Linguistics Graduates, Journal Of Babylon Center For Humanities Studies, July 2026, Volume:16, Issue 7.



This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

المستخلص

تتناول هذه الدراسة تطور الكفاءة النحوية لدى خريجي قسم اللغة الإنجليزية/اللسانيات عند ترجمتهم من العربية إلى الإنجليزية، من خلال التعلم القائم على استخدام المدونات اللغوية. اعتمدت الدراسة على تحليل حالة فردية وتحليل منهجي للأخطاء، حيث تمت مقارنة الأداء النحوي قبل وبعد تدريب محدد باستخدام المدونة. أظهرت النتائج تحسناً ملحوظاً في بنية الجملة، واستخدام الأفعال، وترتيب المحددات. مكّنت أدوات المدونة المتعلمين من تصحيح الأخطاء ذاتياً وإنتاج تراكيب لغوية أكثر طبيعية. وتؤكد الدراسة فعالية التعلم القائم على البيانات في تعزيز الكفاءة النحوية ودقة الترجمة، مع توصيات بدمج المدونات التعليمية في برامج تدريس الترجمة المتقدمة.

Abstract

The purpose of this research is to explore the syntactic growth of graduates of linguistics programs who translate from Arabic into English using corpus-based learning technology. The research investigates

syntactic trends both before and after targeted corpus training by using a case study methodology and a systematic error analysis technique. A significant improvement can be seen in the sentence structure, verb use, and modifier placing, as shown by the results. Through the use of corpus tools, students were able to self-correct and develop structures that were more native-like. The findings of this research provide more evidence that data-driven learning is an effective method for improving syntactic competence and translation accuracy among graduates of linguistics programs. Additional recommendations are made about the use of learner corpora into advanced translation teaching.

1.1 Introduction

In the last few decades, the field of translation studies has used technology more and more to improve both language skills and the quality of translations. Corpus-based learning has become an important method among these new ideas because it uses real language data directly in translator training programs. This research focuses on the syntactic improvements that graduates who have focused in linguistics have obtained via the use of corpus-based learning. More specifically, the study examines how well these graduates do in translations between Arabic and English. The purpose of this project is to offer empirical evidence on the usefulness of corpus-driven education in encouraging syntactic correctness and complexity within translation practice. This will be accomplished via the identification and analysis of syntactic structures.

1.2 Research Objectives

These are the key goals that this research aims to accomplish:

1. Conduct an analysis of the syntactic improvements that graduates of linguistics programs have achieved via the use of corpus-based learning in Arabic-English translations?
2. It is important to conduct a comprehensive error analysis in order to discover common syntactic problems and the possible causes of these errors.
3. When it comes to improving syntactic competence in translation training, it is important to investigate the pedagogical implications of corporation-based learning.

1.3 Research Questions

This research endeavors to provide answers to the following questions:

1. In the field of Arabic-English translation, to what degree does the use of corpus-based learning contribute to the syntactic growth of graduates of linguistics programs?





2. When it comes to their translations, what are the most prevalent grammatical faults that may be encountered?

3. In order to properly address these grammatical issues, what are some ways that corpus-based instructional techniques might be improved?

1.4 Significance of the Study

Through the provision of a syntactic viewpoint that is specifically focused on corpus-based learning outcomes, this study makes a contribution to the area of translation studies. Syntactic gains are an area that is often disregarded in favor of lexical or semantic analysis; this article fills a gap in the current body of work by bringing attention to syntactic gains. The results also have practical implications for educators of translators, curriculum designers, and applied linguists who are working toward the goal of improving syntactic competency using data-driven learning methodologies.

2. Theoretical Framework and Literature Review

2.1 Introduction

The present chapter delineates the theoretical foundations and critically reviews the relevant literature underpinning the current study. It aims to establish a robust conceptual framework by examining theories related to corpus-based learning in translation pedagogy, syntactic analysis in translation, and prior empirical studies investigating the impact of corpus use on syntactic competence.

2.2 Theories of Corpus-Based Learning in Translation Pedagogy

Corpus-predicated knowledge, which follows the principles of data-driven knowledge (DDL), is an important thruway to educate language. preferably of exactly mastering regulations, savants exercise massive collections of real handbooks (called corpora) to discover how language actually works. This path, founded by Johns in 1991, puts savants in the automobilist's seat, allowing them to get for themselves.

In the world of paraphrase, these tools are game-changers. They portray as a reality check, letting savants see how native speakers exercise language. By assaying these corpora, savants can pick up on natural patterns, which helps their paraphrases sound farther authentic and accurate. The main paragon behind this is to promote learner autonomy and contextualized knowledge. It moves beyond the limitations of simple ABC regulations, giving away away savants a deeper understanding of syntax and meaning by showing off off them language in its natural niche.

2.3 Syntactic Analysis and Its Role in Translation

Syntactic competence constitutes a core component of translation proficiency, encompassing knowledge of sentence structure, grammatical relations, and phrase order (Catford, 1965; Newmark, 1988). Precise syntactic transfer from source to target language is pivotal for maintaining meaning and stylistic appropriateness. Errors in syntax can lead to misinterpretation or awkwardness, compromising translation quality (Molina & Hurtado Albir, 2002). Corpus-based approaches offer translators empirical evidence of syntactic patterns, facilitating syntactic disambiguation and improved structural choices (Kenny, 2014).

2.4 Previous Empirical Studies on Corpus Impact on Syntactic Competence

Existing literature on corpus-based translation learning predominantly focuses on lexical and semantic dimensions (Tono, 2014; Zanettin, 2017). However, a growing body of research explores syntactic improvements. For instance, Kennedy (2016) demonstrated that corpus consultation significantly reduced syntactic errors among translation trainees. Similarly, Yoon and Jo (2020) observed enhanced syntactic complexity in student translations after corpus exposure. Nonetheless, few studies explicitly target linguistics graduates or the Arabic-English language pair, highlighting a lacuna addressed by this study.

2.5 An Explanation of What We Mean When We Translate Syntax

The arrangement of words and phrases in order to construct sentences that are accurate from a grammatical standpoint is referred to as syntax. When it comes to translation, one of the most significant challenges is determining how these phrase patterns manifest themselves in two distinct languages. Because of the substantial variances in word order and the manner in which verbs and nouns are required to agree, Arabic and English are especially challenging languages to communicate in. As a consequence of this, translators often encounter substantial difficulties. This is where corpus-based learning comes into play; it provides translators with a method to detect and manage these intricate patterns by making use of relevant data from the actual world.

2.6 A Deficit in the Existing Research

There has not been sufficient research conducted on the specific influence that corpus-based learning has on improving syntax, particularly in certain language pairings. This is despite the fact that the majority of people feel that it is an excellent teaching tool. The subject of translation between Arabic and English is a wonderful example; because to the significant syntactic variations between the two languages, we want more study that is specifically geared to investigate how corpus



technologies truly assist translators in improving their syntactic abilities. This project seeks to make a significant addition to the continuing discourse about how we may utilize corpora to teach better syntax by integrating existing theories with practical data. This will be accomplished by combining this evidence with established theories.

3. Methodology

3.1 Research Design

In order to provide a full discussion of the syntactic earnings that can be accomplished by corpus-based literacy in Arabic-English restatement by graduates of linguistics programs, this research used a mixed-styles exploratory methodology. Both statistical analysis of grammatical faults and an in-depth comprehension of the existential views of actors are made possible by the merging of quantitative and qualitative research methods. In the quantitative part, the careful detection, bracketing, and quantification of syntactic problems in restated textbooks are the primary focuses. It is a coincidence that the qualitative component makes use of a case study technique, which incorporates semi-structured interviews, in order to investigate the actors' habits of discussing the corpus, their approaches to problem-solving, and their opinions on syntactic improvement.

This binary methodological station is crucial in order to reflect the multidimensional nature of restatement literacy, which is characterized by the intersection of quantifiable verbal labors with cognitive and metalinguistic processes. Additionally, the mixed-style approach improves the validity and trustworthiness of results by triangulating data sources and logical methods of doing things throughout the research process.

3.2 Participants

The actors in the research are thirty individuals who have graduated from a reputable academic school in the field of linguistics. These individuals were selected via the use of purposeful slicing to ensure that they are capable and applicable. A comprehensive training program in corpus-based restatement methods has been completed by every actor, and they all maintain intermediate to advanced skill in both Arabic and English. The inclusion criterion required expertise with electronic corpus tools, and it was emphasized that the performance of the actors would represent the true effect of using the corpus rather than the act of being exposed to the original material.

The demographic mix consists of a balanced representation of both genders, as well as an age range that spans from early to late twenties. This cohort is representative of new graduates who are at the beginning

phases of their professional restatement careers. As a result, they provide insight on corpus-based literacy commodities that are in the process of being developed.

3.3 Data Collection Instruments

Multiple reciprocal instruments were used throughout the data gathering process. These instruments were meant to capture both objective verbal performance and private existential accounts. In the beginning, the members of the cast participated in a controlled restatement assignment, which consisted of translating a number of Arabic textbooks into English. These textbooks were meticulously selected to carefully contain a wide range of syntactic structures, ranging from straightforward declarative rules to intricate clauses that are embedded inside sentences. Because of this, it was guaranteed that the syntactic problems were both complete and representative.

In the second step of the process, a methodology for mistake analysis was meticulously designed. This procedure was based on conventional syntactic ordering that were appropriate to Arabic-English restatement. Errors such as subject-verb agreement mismatches, erroneous word order, improper use of relative clauses, and expression structure diversions were able to be coded in a systematic manner thanks to this procedure. To ensure the reliability of the quarantine-rater, the coding was carried out independently by two knowledgeable raters.

Post-translation, semi-structured interviews were carried out in order to get insight into the activities that actors took with regard to the utilization of the corpus. The purpose of these interviews was to investigate the methods by which actors accessed corpus material, the methods that were used to resolve syntactic inscrutability, and the actors' perceptions of the effectiveness of corpus tools in boosting syntactic delicacy.

3.4 Research Procedure

The exploration was carried out in a series of successive stages that were aimed to preserve the methodological rigor and depth of the investigation. In the beginning, actors participated in a factory training session that focused on advanced corpus discussion methods and syntactic proposition material to Arabic-English restatement. A lively birth of corpus faculty was ensured as a result of this.

After this, the actors worked together to finish the restatement job under controlled settings. This was done in order to exclude extraneous factors such as time pressure or cooperative effect. A digital collection of restatements was made for the purpose of posterior analysis.

In addition, the procedure of mistake rendering was used for the corpus of restatements. Syntactic mistakes were deciphered after being





enciphered by two independent raters in accordance with the stated process.

3.5 Data Analysis

In order to conduct an analysis of our data, we merged narratives and numerical data.

In the beginning of the quantitative analysis, we made use of descriptive statistics to determine the frequency of syntactic mistakes as well as the locations where they were found. Then, we used comparative analysis to determine whether or not there were any patterns that indicated a connection between the utilization of corpora and the enhancement of syntax.

The journey did not end there. We thoroughly examined the interview transcripts in order to do the qualitative analysis. Our goal was to identify recurring trends about the manner in which the participants engaged with the corpora, their understanding of syntax, and the translation procedures they used.

By combining the numerical data with the narratives, we were able to get a far more profound comprehension of the efficacy of corpus-based learning in terms of enhancing the syntactic abilities of these grads. Because of this dual method, our results were significantly reinforced, and as a result, they were more relevant and applicable.

4. Analysis and Discussion

4.1 Overview of Syntactic Errors Identified

This chapter presents a detailed analysis of the syntactic errors linked in the Arabic- English restatements produced by linguistics graduates, alongside an interpretive discussion of the results. The error analysis is framed within the environment of corpus- grounded literacy, examining to what extent corpus discussion contributed to syntactic advancements.

The analysis distributed syntactic errors into several crucial types, videlicet subject- verb agreement, word order violations, abuse of relative clauses, and incorrect expression structure. The frequency and distribution of these errors were quantitatively assessed and are epitomized in Table 4.1.

Error Type	Frequency	Percentage (%)
Subject-Verb Agreement	45	22.5
Word Order Violations	38	19
Misuse of Relative Clauses	30	15
Incorrect Phrase Structure	57	28.5

Error Type	Frequency	Percentage (%)
Other Minor Syntactic Errors	30	15
Total	200	100

Table 4.1: Distribution of Syntactic Errors in Translations

It was determined via the analysis of the data that the most prevalent error was an erroneous sentence construction, which accounted for 28.5% of all mistakes. According to past research, which emphasizes how difficult it is to translate complicated statements from one language to another, this statement is consistent. One further thing that we discovered was that subject-verb agreement was a substantial issue, which indicates that there are continued difficulties with morphosyntactic consistency.

4.2 Impact of Corpus-Based Learning on Error Reduction

An examination of the errors that were committed by the participants before and after they used the corpus revealed a discernible decrease in a variety of errors. An illustration of the shift in mistake frequency for both translation jobs can be seen in Figure 4.1.

Error Type	Pre-Corpus Frequency	Post-Corpus Frequency	Reduction (%)
Subject-Verb Agreement	60	45	25
Word Order Violations	52	38	26.9
Misuse of Relative Clauses	42	30	28.6
Incorrect Phrase Structure	70	57	18.6
Other Minor Syntactic Errors	40	30	25
Total	264	200	24.2

Table 4.2: Comparison of Syntactic Errors Pre- and Post-Corpus Consultation

The data shows that after using a corpus, participants reduced their overall errors by 24.2%. This highlights just how effective corpus-based learning is at improving syntactic accuracy. We saw the most significant improvements in areas that are typically tough for Arabic-English



translators, like relative clauses and word order, which are structurally very different between the two languages.

4.3 Qualitative Insights from Participants' Reflections

The interviews provided even more depth, showing us exactly how using a corpus affected the participants' syntactic performance. They consistently said that corpora helped them clear up confusion, especially in complex sentences. For example, one participant noted, "Using the corpus examples helped me see the authentic patterns for using a relative clause, which I always found confusing in textbooks."

Additionally, several participants stressed how these tools helped them build "syntactic intuition," allowing them to double-check their own work against real-world data. This supports Baker's (2020) argument that using corpora helps bridge the gap between theoretical knowledge and practical application.

4.4 Discussion

Combining our quantitative and qualitative data clearly shows that corpus-based learning is the main reason these linguistics graduates became more proficient in Arabic-English translation. The significant reduction in errors, especially in tricky areas like relative clauses and word order, confirms that consulting a corpus is particularly helpful for solving these syntactic issues.

In a similar vein, the actors' observations highlight the fact that corpus tools serve not just as error-correction sources but also as devices for attaining a more profound syntactic awareness. This binary function improves both the subtlety and the ignorance of restatements, which in turn contributes to the development of learner autonomy and confidence.

The permanence of some mistakes, particularly in expression structure, shows that corpus-based literacy should be reinforced by specific syntactic teaching and continuing practice. This is despite the fact that these earnings have been made. In light of these results, integrative educational models are recommended. These approaches integrate corpus tools with unambiguous alphabet instruction that is adapted to multilingual syntactic differences.

5. Conclusion and Recommendations

5.1 Conclusion

The purpose of this research was to investigate the influence that corpus-based literacy has on the syntactic performance of Arabic-English restatements that are created by graduates of linguistics training programs. The purpose of this investigation was to determine whether or not exposure to actual verbal material via corpus discussion may improve

syntactic skill in restatement tasks. The investigation used a binary methodological approach, which consisted of quantitative error analysis and qualitative reflection.

The results indicate that there is a significant decrease in the number of significant syntactic mistakes that occur after the corpus is engaged, notably in the domains of subject-verb agreement, relative clause operation, and word order. Through qualitative perceptivity, actors were able to express enhanced syntactic awareness and confidence while working with corpus tools, which provided support for the quantitative findings that were obtained.

The research substantiates the hypothesis that corpus-based literacy enhances deeper syntactic processing and supports data-driven restatement methods. This finding is the most astounding of all the findings. Corpora are tools that bridge the gap between abstract grammar principles and practical restatement prosecution. They do this by introducing students to language patterns that may be encountered in the real world. In addition to correcting individual problems, comparable technologies also help the development of syntactic suspicion and stylistic excellence over the course of a longer period of time.

Nevertheless, the research showed that despite the discussion of the corpus, many grammatical issues, particularly those associated with complicated expression patterns, continue to exist. A conclusion that may be drawn from this is that while corpus tools are essential, they are not a fix. In order to achieve successful integration, an organized educational landscape, assistance from teachers, and reflective practice are all necessary components.

5.2 Recommendations

In light of the results, a number of suggestions that are educational and exploration-based have been made.

5.2.1 Pedagogical Recommendations

- Incorporating Corpus Tools into the Curriculum Academic restatement programs should include knowledge of corpus into their curricula. This will ensure that students are knowledgeable not just about how to use corpus tools, but also about how to critically evaluate corpus data.

- Syntactic-concentrated Corpus Tasks Preceptors should construct restatement exercises that clearly target syntactic structures that are prone to mistake. These structures include bedded clauses, verb-preposition phrases, and the complexity of noun expressions.

- Blended training Models Literacy training that is based on the corpus should be supplemented with conventional alphabet instruction. This will allow students to integrate syntactic principles onto empirical examples.



- Learners should be encouraged to reflect on their restatement processes, specifically comparing their pre-corpus labors to their post-corpus labors in order to identify patterns of improvement and areas that need more foundation.

5.2.2 Research Recommendations

- Longitudinal Studies Further investigation should use a longitudinal design in order to investigate the long-term influence of corpus-grounded training on syntactic development over the course of several semesters or academic periods.

Why Cross-Disciplinary relative Studies research that involves researchers from departments of literature, linguistics, and restatement has the potential to give insights into the syntactic trends and corpus operation efficacy that are distinctive to each field.

- Corpus kinds and stripes: Additional research should investigate the influence of various corpus kinds (for example, learner corpora vs professional corpora) or stripes (for example, news, legal, and erudite) on syntactic delicacy in restatement.

5.3 Final Remarks

This study makes a contribution to the expanding body of academic literature that argues that corpus-based learning is an efficient method for improving translation quality, particularly in syntactic areas. The research revealed how corpus tools facilitate the operationalization of theoretical linguistic knowledge in realistic translation situations by focusing the case study on persons who had graduated from linguistics programs and had theoretical linguistic knowledge. As a result of these results, there is an urgent need for a change in educational practices toward evidence-based translation training that places equal importance on language precision and analytical independence.

In conclusion, while there are certain restrictions associated with corpus-based learning, it is an important step forward in the field of translation education. It is possible for it to serve as a cornerstone for the cultivation of syntactically competent and critically aware translators who are capable of negotiating the complexity of Arabic-English syntactic transfer if it is integrated thoughtfully and pedagogical assistance is provided.

References

- Al-Shboul, Y. (2017). Error analysis of Arabic–English translation: A study of syntactic and lexical problems. *Arab World English Journal*, 8(1), 201–218. <https://doi.org/10.24093/awej/vol8no1.15>
- Baker, M. (2020). *Translation and conflict: A narrative account* (2nd ed.). Routledge. <https://doi.org/10.4324/9780429432073>
- Catford, J. C. (1965). *A linguistic theory of translation: An essay in applied linguistics*. Oxford University Press.
- Chomsky, N. (1957). *Syntactic structures*. Mouton.
- Hussein, A., & Al-Khalil, M. (2019). Syntactic interference in Arabic–English translations by students. *International Journal of Applied Linguistics & English Literature*, 8(2), 45–52. <https://doi.org/10.7575/aiac.ijalel.v.8n.2p.45>
- Kenny, D. (2014). *Corpora in translation studies*. Routledge. <https://doi.org/10.4324/9781315759951>
- Kennedy, G. (2016). *An introduction to corpus linguistics* (2nd ed.). Routledge. <https://doi.org/10.4324/9781315842028>
- Molina, L., & Hurtado Albir, A. (2002). Translation techniques revisited: A dynamic and functionalist approach. *Meta: Journal des Traducteurs*, 47(4), 498–512. <https://doi.org/10.7202/008033ar>
- Newmark, P. (1988). *A textbook of translation*. Prentice Hall.
- Olohan, M. (2017). *Introducing corpora in translation studies* (2nd ed.). Routledge. <https://doi.org/10.4324/9781315694375>
- Smith, J. (2021). Learner corpora in translator training: A syntactic perspective. *Language and Translation*, 33(2), 105–123.
- Tono, Y. (2014). Learner corpora: Design, development and applications. In S. Granger, G. Gilquin, & F. Meunier (Eds.), *The Cambridge handbook of learner corpus research* (pp. 101–123). Cambridge University Press. <https://doi.org/10.1017/CBO9781139649414.007>
- Vinay, J. P., & Darbelnet, J. (1995). *Comparative stylistics of French and English: A methodology for translation* (J. C. Sager & M.-J. Hamel, Trans.). John Benjamins. (Original work published 1958). <https://doi.org/10.1075/btl.11>
- Yoon, H., & Jo, J. (2020). Using learner corpora for error analysis in translation training. *Linguistics and Education*, 60, 100862. <https://doi.org/10.1016/j.linged.2020.100862>
- Zanettin, F. (2017). *Translation-driven corpora: Corpus resources for descriptive and applied translation studies* (2nd ed.). Routledge. <https://doi.org/10.4324/9781315760902>

Appendix A: Sample Translations and Error Analysis

Sample	1	–	Original	Arabic	Sentence:
Student	Translation	(Before	Corpus	Exposure):	
Revised	Translation	(After	Corpus-Based	Training):	
Syntactic Errors Detected:					
•Missing auxiliary verb (is)					



Analyzing Syntactic Gains through Corpus-Based Learning: A Case Study and Error Analysis of Arabic-English Translations by Linguistics Graduates

- Incorrect lexical choice (*level* instead of *quality*)
- Unnecessary article (*the* before *rural areas*)

Sample 2 – **Original Arabic Sentence:**
 "إن تطویر البنية التحتية أمر ضروري للنمو والاقتصادی"

Student Translation (Before Corpus Exposure):
 Develop the infrastructure is important for the economic growth.

Revised Translation (After Corpus-Based Training):
 Developing infrastructure is essential for economic growth.

Syntactic Errors Detected:

- Incorrect verb form (*develop*)
- Article misuse (*the economic growth*)
- Lexical enhancement after corpus usage (*essential* vs. *important*)

Sample 3 – **Original Arabic Sentence:**
 "تمت المصادقة على الاتفاقية من قبل البرلمان بالإجماع"

Student Translation (Before Corpus Exposure):
 The agreement has been certified by the parliament unanimously.

Revised Translation (After Corpus-Based Training):
 The agreement was unanimously ratified by the parliament.

Syntactic Errors Detected:

- Wordiness and awkward structure
- Misuse of passive form
- Inappropriate verb choice (*certified* instead of *ratified*)