



Castledown

 OPEN ACCESS

Australian Journal of Applied Linguistics

ISSN 2209-0959

<https://www.castledown.com/journals/ajal/>

Australian Journal of Applied Linguistics, 6(3), 205–218 (2023)

<https://doi.org/10.29140/ajal.v6n3.1092>

The Role of Corpora in Enhancing Translation Accuracy and Fluency Feasibility of Using Corpora as a Tool in Translation Practice



NOURELDIN MOHAMED ABDELAAL

*Faculty of Languages
October University for Modern Sciences and Arts (MSA),
nmnedlin@msa.edu.eg
nourabdelal@yahoo.com*

Abstract

Translators employ an array of tools to streamline the translation process, ensuring precision and coherence. The necessity of these tools is particularly pronounced when tackling authentic materials, such as texts from the United Nations (UN). Among these tools, corpora stand out as a pivotal resource. In light of this, this study undertook a comprehensive qualitative and quantitative exploration into the impact of UN parallel corpora on novice translators. The research sought to address two fundamental questions: Does the utilization of parallel corpora significantly enhance translation performance? Furthermore, what specific areas of improvement, if any, are facilitated by the incorporation of parallel corpora?

To address these questions, a cohort of 30 students undertook two translation tasks of equal complexity. The initial task was conducted without the support of UN corpora, while the subsequent task integrated the use of these corpora. Employing a t-test, the study scrutinized the statistical significance of divergence among the participants. The outcomes illuminate that the integration of corpora within translation pedagogy, particularly for UN texts, plays an indispensable role. This integration furnishes students with authentic source materials, facilitating comprehension of intricate terminologies inherent in these documents. Furthermore, corpora enable students to discern underlying linguistic patterns, encompassing syntax, collocations, and discourse structures. This linguistic insight equips translators to faithfully convey intended meanings and registers in their translations. Notably, the T-test outcomes establish a substantial positive influence of corpora utilization on the achievement of students.

Keywords: UN corpora; Sketch Engine; t-test; translation performance.

Copyright: © 2023 Noureldin Mohamed Abdelaal. This is an open access article distributed under the terms of the Creative Commons Attribution Non-Commercial 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. **Data Availability Statement:** All relevant data are within this paper.

Introduction

Translator Tools

Professional translators routinely leverage an array of tools, with terminology management software, Computer-Assisted Translation (CAT) tools, and corpora standing out as prominent examples. These tools not only augment translator productivity but also fortify consistency, consequently yielding reductions in translation expenditures (Zampieri and Vela, 2014). Among these tools, corpora are particularly noteworthy, encompassing distinct categories, such as comparable corpora, parallel corpora, and monolingual corpora. Within this context, a corpus can be defined as an assembly of naturally occurring machine-readable texts (Hu, 2016), thereby possessing certain inherent attributes.

The first defining attribute is ‘representativeness’. A corpus must encompass a composition and distribution of materials that accurately mirror a specific language variety (Hu, 2016). A pertinent exemplar is the United Nations (UN) corpus, which faithfully captures the formal language employed within the UN domain. The second hallmark of corpora, as elucidated by Hu, is digitization. This facet facilitates seamless exploration and navigation of the corpus using an assortment of specialized corpus tools. With these fundamental corpus characteristics delineated, the subsequent discourse delves into the distinct types of corpora within the realm of translation studies.

Types of Corpora in Translation Studies

Hu (2016) identifies four distinct types of corpora within the realm of translation studies: comparable corpora, translational corpora, interpreting corpora, and parallel corpora. Comparable corpora have proven instrumental in investigating matters of equivalence and shedding light on the intricate challenges of non-equivalence encountered by professional translators. These corpora have also proven to be a valuable pedagogical resource for translation instructors (Zanettin et al., 2003; Beeby et al., 2009; Frankenberg-Garcia, 2016).

Monolingual corpora, on the other hand, serve as a lens through which to observe language usage, word sequences, and collocations. They have been harnessed to unravel the subtle distinctions in meaning between closely related synonyms.

Parallel corpora, as articulated by Baker (1995), comprise source texts (STs) paired with their corresponding target translations. These corpora can assume bilingual or multilingual forms and can be classified as unidirectional, bidirectional, or multidirectional as outlined by Hu (2016). Unidirectional parallel corpora consist of STs in one language and their translations in another. In the case of bidirectional parallel corpora, STs are juxtaposed with translations in two languages. An instance of this can be found in the UN corpus, where STs in English are translated into Arabic, and reciprocally, STs in Arabic are translated into English. Multidirectional parallel corpora entail STs in one language aligned with their translations in two or more languages.

The utility of parallel corpora spans diverse applications. These corpora have played a pivotal role in training translation systems, examining the translation strategies employed by professionals in tackling intricate discourse, and unraveling nuances in translation shifts and divergences (Johansson & Hofland, 2000; Johansson, 2007; Xiao & Dai, 2014; Frankenberg-Garcia, 2014, 2016).

Larger corpora extend myriad pedagogical avenues for translation instruction, translator education, and professional advancement (Alhassan et al., 2021). Research into corpora in translation (e.g., Alhassan,

2015; Fantinuoli & Zanettin, 2015) underscores the potential advantages of employing parallel corpora, whether bilingual or multilingual, within the translation classroom. Corpora hold significant value in translation endeavors, exposing students to a diverse array of parallel texts. Moreover, they serve as a cornerstone in the training data for statistical machine translation and the development of bilingual lexicons and terminologies (Sabtan, 2016). Their efficacy is further pronounced in addressing translation challenges, maintaining terminological consistency, offering translation alternatives in varying contexts, and facilitating idiomatic expression. These repositories expedite translation processes, offering insights into the strategies adopted by seasoned translators across different types of translation (Alhassan, 2015).

Alignment of Texts

Text alignment is a pivotal process that can be executed at varying granularities, such as the sentence or paragraph level. While alignment at smaller units, like sentences, holds greater significance for translation applications, achieving alignment at this level is notably more intricate. This challenge is particularly pronounced in language pairs, such as English and Arabic. In the realm of sentence-level alignment, two distinct techniques commonly emerge: sentence-length-based alignment and lexis-based alignment.

The sentence-length-based technique, originally proposed by Brown et al. (1991) and further advanced by Gale and Church (1994), posits a statistical correlation between the lengths of source text (ST) sentences and their corresponding target text (TT) sentences. In essence, this method identifies sentence matches by quantifying the number of words or characters in both the ST and TT and subsequently subjecting these metrics to mathematical scrutiny to ascertain alignment.

Conversely, the lexis-based approach involves alignment at the lexical level, focusing on individual words (Kay & Roscheisen, 1993). While lexis-based alignment boasts heightened accuracy, it inherently demands a substantial investment of time and effort. In essence, text alignment is a multifaceted endeavor with implications that span translation practices. The choice between these alignment techniques necessitates a trade-off between precision and efficiency, whereby sentence-length-based alignment emphasizes expediency and lexis-based alignment underscores meticulousness.

Challenges of Using Corpora

While the integration of corpora in translation practices presents numerous advantages, there are distinct challenges that underscore the complexity of their implementation. One pivotal challenge pertains to the intricacies of corpus design, which demand meticulous attention to factors like representativeness, typicality, and scale (Kennedy, 2014). Ensuring that a corpus embodies these qualities can be a formidable task.

Additionally, a noteworthy challenge emerges from the characteristics of the Arabic language itself. In many instances, Arabic is not readily machine-readable, thereby complicating the process of corpus analysis. The intricacies of the Arabic language further amplify these challenges. Accessibility is another constraint. Many corpora, particularly high-quality and specialized ones, are not available free of charge. This aspect can hinder their widespread adoption, rendering them less popular as tools in the translation process and even in second language (L2) classrooms. Moreover, the tools devised for exploring parallel corpora often lack user-friendliness, being tailored primarily to cater to researchers rather than language learners (Doval & Nieto, 2019, p. 10). This design oversight can create barriers for individuals seeking to harness corpora effectively.

Furthermore, a notable challenge lies in the limited awareness surrounding the importance and applications of corpora among instructors. This lack of awareness, coupled with the dearth of accessible tools for accessing parallel corpora, poses a notable obstacle to the broader utilization of corpora in various educational contexts.

In summation, the incorporation of corpora into translation endeavors faces a spectrum of challenges. These range from the intricacies of corpus design and the unique characteristics of certain languages, such as Arabic, to issues of accessibility, user-friendliness, and awareness among instructors and learners. Overcoming these challenges is pivotal for harnessing the full potential of corpora in the realm of translation and language education.

Applications of Parallel Corpora

Parallel corpora have been harnessed in diverse ways, each shedding light on the intricate nuances of language usage and aiding various translation-related processes. One significant application lies in the linguistic and stylistic comparison of two languages. Løken (1997), for instance, utilized parallel corpora to juxtapose the employment of possibility expressions in English and Norwegian. His analysis revealed substantial differences between the languages, particularly in terms of cognitive possibility expressions. Furthermore, he noted that Norwegian adverbs and constructions involving modal verbs and adverbs contribute to the translation of English epistemic modal verbs. Similarly, Altenberg (1998) employed English-Swedish corpora to elucidate both commonalities and disparities in conjunctives and sentence openings between the two languages. Notably, he discovered distinctions in the manner sentences are initiated, with Swedish not invariably commencing with a subject.

A second pivotal use of parallel corpora is in the creation of dictionaries. An illustrative instance is the *Collins COBUILD* dictionary, which is underpinned by the Sketch Engine and takes a corpus-based approach. By utilizing corpora, this dictionary compiles source texts along with their corresponding translations. This proves invaluable for tackling challenging aspects like idiomatic expressions and collocations, aiding both novice and experienced translators. Additionally, corpora serve as a well-spring of examples that illuminate natural language usage.

Machine translation constitutes another significant arena where parallel corpora play an essential role. Machine translation encompasses both translation systems, exemplified by Google Translate, and CAT incorporating machine translation functionalities. The evolution of machine translation has traversed diverse stages, from rule-based approaches to the prevailing hybrid models. Rule-based machine translation (RMBT) involves morphological, syntactical, and semantic analysis of both the source text and target text, utilizing parsers, analyzers, and dictionaries (Hu, 2016).

Statistical machine translation (SMT), an alternative approach, employs statistical techniques and bilingual corpora to facilitate translation. Example-based machine translation relies on analogy, where the translation of a given sentence or word is determined by drawing parallels from extensive corpora. The Neural Machine Translation (NMT) approach leverages neural systems for translation prediction, harnessing artificial intelligence. Finally, hybrid machine translation combines these various techniques, with all except RMBT relying extensively on corpora (Hu, 2016).

Use of Parallel Corpora in Translation-Related Research

A parallel corpus serves as a potent instrument for delving into translation strategies, methods, translational norms, and language equivalence, particularly at the lexical and sentential levels. Such investigation is rooted in the analysis of aligned source texts (STs) and target texts (TTs) (Hu, 2016, p. 41).

For instance, Qin and Wang (2004) undertook a comparison between the English word ‘dog’ and its Chinese translation, unearthing the prevalent utilization of the foreignization strategy.

Moreover, a parallel corpus proves invaluable in the identification of distinct translator styles. This entails an examination of variables such as the type/token ratio, word frequency, average sentence length, and the occurrence frequency of specific lexicons or syntactic structures in the studied texts.

Corpora also serve as a valuable resource for contrasting STs and their translations in terms of lexical density and the prevalence of diverse parts of speech. An illustrative case is the study conducted by Hu and Zou (2009), wherein they contrasted a Shakespearean play’s ST with two distinct translations by Liang Shiqiu and Zhu Shenghao. The comparison encompassed metrics such as the type/token ratio, average sentence length, frequency of particular words, and syntactic structures. This investigation unveiled noteworthy disparities: Zhu Shenghao’s translation exhibited a higher total token count and average sentence length compared to Liang Shiqiu’s, while also displaying a lower type/token ratio. The study further revealed Liang’s predilection for adopting foreignization as a translation approach.

In essence, a parallel corpus serves as a versatile analytical tool, empowering researchers to probe diverse facets of translation, including strategies, stylistic nuances, equivalence assessment, and linguistic characteristics. The insightful comparisons and contrasts drawn from these analyses provide valuable insights into translation dynamics and strategies.

Comparable Corpora

Comparable corpora, as distinct from parallel corpora, encompass two sets of texts that are not direct translations of each other. These corpora can be categorized as monolingual, bilingual, or multilingual collections. In the case of a monolingual corpus, it amalgamates both non-translated texts and their translated counterparts within the same language (Hu, 2016). Baker (1995) conceptualizes a monolingual comparable corpus as comprising two distinct text sets: the first encompasses original language texts, while the second contains translations. For instance, one could compile an English news corpus centered on a topic like climate change and concurrently assemble another English news corpus comprised of translations of Arabic news into English. These two collections would constitute a monolingual comparable corpus.

Alternatively, a bilingual or multilingual comparable corpus consists of texts across multiple languages that do not share a direct translation relationship (Hu, 2016). In essence, these texts are chosen for comparison due to thematic or contextual similarities rather than being translations of one another. This makes such corpora a valuable resource for exploring linguistic and cultural variations across languages without the constraints of translation alignment.

Applications of Comparable Corpora

As delineated by Hu (2016), comparable corpora serve diverse functions, offering a versatile toolset for linguistic analysis and translation studies. Several of these applications include:

- **Contrastive Analysis:** Comparable corpora can be employed to conduct a contrastive analysis between two languages. For instance, researchers might investigate the usage of discourse markers by native and non-native writers of English, revealing insights into language proficiency and linguistic variations.
- **Comparison of Translated and Non-Translated Texts:** Comparable corpora enable the examination of similarities and differences between translated and non-translated

texts, encompassing lexical, sentential, semantic, and textual levels. This affords a nuanced understanding of the impact of translation on various linguistic aspects.

- **Investigation of Translation Tendencies:** Comparable corpora facilitate investigations into translation tendencies, including factors such as a translator's personality, gender, and style. These inquiries contribute to a deeper comprehension of the human elements shaping translation processes.
- **Exploration of Translation Norms:** Comparable corpora offer a means to delve into translation norms. For instance, Laviosa (1996) conducted an examination of news reports within a comparable corpus, analyzing metrics such as type/token ratio, lexical density, and average sentence length. Her findings indicated a tendency towards simplification and normalization, aligning with translational universal norms as posited by Baker.

In essence, comparable corpora empower researchers to explore a broad spectrum of linguistic and translation-related phenomena, from language contrasts to the influence of translation on textual characteristics and normative tendencies.

Previous Studies

Numerous studies have harnessed corpora to delve into grammatical, syntactic, and lexical intricacies within translation. Johansson and Hofland (2000) delved into the utilization of modal auxiliaries and their contextual nuances. Blanco (2009) scrutinized the realm of adverbials, while Frankenberg-Garcia (2016) directed attention towards lexical considerations, particularly investigating loan words and their usage.

Despite the existence of a substantial body of theoretical research (e.g., Alhassan, 2015; Alotaibi, 2016; Alotaibi, 2017) underscoring the significance of incorporating parallel corpora into translation pedagogy, the empirical landscape remains unexplored. Notably, no empirical study has been conducted to ascertain the efficacy of parallel corpora utilization. Thus, this present research undertakes the task of investigating the application of parallel corpora within a specialized translation class, with the explicit aim of enhancing translation quality.

By addressing this gap in empirical research, the study seeks to contribute substantively to the discourse on translation education, offering empirical insights into the practical implications of utilizing parallel corpora for skill enhancement in translation classrooms.

Research Questions

- Is there a significant positive impact of the use of parallel corpora on improving translation performance?
- What areas of improvement (if any) does the use of parallel corpora help in?

Methodology

Thirty students were assigned a translation task to translate a text using Sketch Engine as a tool in translation to search for collocations and new words in the UN parallel corpora. They were taking the UN Translation course, and they were asked to complete two tasks as follows:

Task 1: Translate a 200-word UN document in class. They were allowed to use dictionaries.

Task 2: Translate a 200-word UN document in class using the UN parallel corpora available at Sketch Engine.

The two tasks are of the same genre and at the same linguistic difficulty. Lexile was used to ensure the level of difficulty quantitatively, while the researcher's experience was used to ensure the level of difficulty qualitatively. The two tasks were marked by the researcher using a rubric. T-test was run to check whether there is a significant difference between the students' performance in the two given translation tasks. The researcher examined the translations to see qualitatively the differences in performance between the two translation tasks among the translation students, and the areas of improvement.

Results

Qualitative Results

As explained in the methodology section, the participants were asked to translate some UN texts without using the corpus. After that, they were asked to redo the same translation task using the UN corpora. The results of the study can be summarized in Table 1 below.

Table 1 *Qualitative results*

Term/Expression/Phrase	Without Corpus	Type of Error	With Corpus
1. Non-observed economy	اقتصاد غير مراقب اقتصاد غير ملاحظ اقتصاد غير مرصود	Unidiomatic translation Not collocational	الاقتصاد غير المشمول بالاحصاءات الاقتصاد غير المشمول بالاحصاء الاقتصاد غير الخاضع للمراقبة الاقتصاد غير الملحوظ
2. Secretariat	السكرتارية	Unidiomatic/ inaccurate	الأمانة العامة
3. Electronic consultation	استشارة إلكترونية	Unidiomatic	التشاور بالوسائل الإلكترونية مشاورات الكترونية
4. Feedback	تعليق إفادة	Unidiomatic	تغذية مرتدة تغذية استرجاعية تغذية عكسية
5. Updated	حديثة منطورة	Unidiomatic	مستكملة
6. In-depth review	مراجعة متعمقة مراجعة عميقة	Unidiomatic	استعراض متعمق
7. Macroeconomic frameworks	أطر الاقتصاد الكلي أطر الاقتصاد الكبير	Sometimes unidiomatic	أطر الاقتصاد الكلي أطر اقتصاد كلية أطر عمل اقتصادية شاملة
8. Pursuant to its resolution	طبقاً لقرارها	Unidiomatic	عملاً بالقرار
9. Taking note of	يلاحظ	Unidiomatic/ inaccurate	تحيط علماً بـ
10. Recalling	يستدعي	Unidiomatic/ inaccurate	إذ تشير إلى
11. Panel discussion	نقاش الخبراء نقاش اللجنة	Unidiomatic/ inaccurate	حلقة النقاش
12. The panel discussion was chaired by the President of the Human Rights Council.	ترأست نقاش اللجنة من قبل رئيس مجلس حقوق الإنسان	Inaccurate syntactic pattern	وترأس حلقة النقاش رئيس مجلس حقوق الإنسان.

Term/Expression/Phrase	Without Corpus	Type of Error	With Corpus
13. Adverse impact	التأثير المضاد	Not collocational	التأثير السلبي
14. The United Nations High Commissioner for Human Rights	المفوض السامي للأمم المتحدة لحقوق الإنسان	Inaccurate order of recursive words	مفوض الأمم المتحدة السامي لحقوق الإنسان
15. People in vulnerable situation	على الناس في وضع ضعيف	Not collocational	الأشخاص المعرضين للخطر
16. It was opened by a statement by the United Nations High Commissioner for Human Rights, followed by a video screening of testimonies about the adverse impact of climate change on people in vulnerable situation.	و تم افتتاحه ببيان من المفوض السامي للأمم المتحدة لحقوق الإنسان ، تلاه فحص بالفيديو لشهادات حول التأثير المضاد لتغير المناخ على الناس في وضع ضعيف	Unidiomatic style	وافتح المنتدى ببيان صادر عن مفوض الأمم المتحدة السامي لحقوق الإنسان، تلاه عرض بالفيديو لشهادات حول التأثير السلبي لتغير المناخ على الأشخاص /المعرضين للخطر الأشخاص الذين هم في أوضاع هشة
17. The full and effective enjoyment of human rights	الاستمتاع الكامل والفعال بحقوق الإنسان	Not collocational	التمتع الكامل والفعال بحقوق الإنسان
18. The promotion and protection of the rights	تعزيز و حماية حقوق	Unidiomatic	تعزيز حماية حقوق
19. The panel discussion provided an opportunity for States, international organizations and other relevant stakeholders to discuss the adverse impact of climate change on the full and effective enjoyment of human rights by people in vulnerable situations, as well as best practices and lessons learned in the promotion and protection of the rights of such persons.	أتاحت نقاش الحلقة فرصة للدول والمنظمات الدولية وأصحاب المصلحة المعنيين الآخرين لمناقشة الأثر السلبي لتغير المناخ على الاستمتاع الكامل والفعال بحقوق الإنسان من قبل الأشخاص الذين هم في أوضاع هشة ، وكذلك أفضل الممارسات والدروس المتعلمة من تعزيز و حماية حقوق هؤلاء الأشخاص.	Inaccurate syntactic pattern Not collocational	أتاحت حلقة النقاش فرصة للدول والمنظمات الدولية وأصحاب المصلحة المعنيين الآخرين لمناقشة الأثر السلبي لتغير المناخ على التمتع الكامل والفعال بحقوق الإنسان من قبل الأشخاص الذين هم في أوضاع هشة ، وكذلك أفضل الممارسات والدروس المستفادة في تعزيز حماية حقوق هؤلاء الأشخاص.
20. The panel discussion was made accessible to persons with disabilities through the provision of International Sign interpretation and real-time captioning.	تم جعل حلقة النقاش في متناول الأشخاص ذوي الإعاقة من خلال توفير الترجمة الفورية للعلامات الدولية والشرح في الوقت الفعلي.	Inaccurate Literal translation	تم جعل حلقة النقاش في متناول الأشخاص ذوي الإعاقة من خلال توفير الترجمة الفورية بلغة الإشارة و الترجمة النصية المباشرة.

As can be seen in Table 1, there were clear improvements in the translation task when the UN corpora was used by the participants. Improvements were found in:

Maintaining Idiomatic Translation

Maintaining idiomaticity is a notable challenge when translating UN documents, and this concern was observed among novice translators. This issue becomes particularly evident when examining specific examples of translation attempts before and after utilizing the corpora:

- Example 1: Before consulting the corpora, the rendering of ‘non-observed economy’ by participants yielded variations like *اقتصاد غير مراقب*, *اقتصاد غير ملاحظ*, or *اقتصاد غير مرصود*. These initial attempts were unidiomatic. However, with exposure to the UN corpora, the translations were rectified to *الاقتصاد غير الخاضع للمراقبة* or *الاقتصاد غير المشمول بالاحصاء*, underscoring the significance of specialized corpora, especially for novice translators. This also underscores that accurate translations might require substantial expertise within the specific field.
- Example 2: The term ‘Secretariat’ was initially translated as *السكرتارية* by some students, which, while semantically correct, lacks idiomaticity. Upon exposure to the UN corpora, the translation evolved to *الأمانة العامة*, which is not only denotatively accurate but also culturally fitting. This pattern held true for other examples (3–10 and 19) where corpora-guided translations yielded improved idiomaticity.
- Example 4: In rendering ‘feedback’, initial attempts included *إفادة* and *تعليق*, though the latter is more contextually appropriate. However, upon consultation of the corpora, translations were adjusted to *تغذية أسترجاعية*. This case illustrates how direct translations can sometimes appear odd, even if technically correct, and why exposure to corpora assists in achieving idiomatic expression.
- Example 5: The word ‘updated’ saw translations like *منظورة* and *حديثه* initially. However, these translations fell short in capturing idiomatic usage. The corpora provided the insight to translate it as *مستكملة*, reflecting both accuracy and idiomaticity.
- Example 6: The phrase ‘in-depth review’ underwent transformation from *مراجعة متعمقة* to *أستعراض متعمق* after utilizing the corpora. This change indicates the influence of idiomatic usage gleaned from the corpus.
- Example 16: In this case, adherence to the style of the source text (ST) resulted in unidiomatic translations. The importance of considering the target text (TT) style for idiomaticity is evident here.

These examples underscore how specialized corpora, like those from the UN, prove instrumental in aiding translators to not only achieve semantic accuracy but also to capture idiomatic expressions and nuances specific to the target language and context.

Use of Collocations, Terminology and Syntactic Patterns Including Recursive Words

The utilization of corpora proved advantageous in addressing several key challenges in translation, particularly in terms of maintaining collocational accuracy, syntactic patterns, and overall translation precision. The examples provided highlight these benefits:

Collocational and Syntactic Accuracy:

In example 11, participants faced challenges due to the syntactic structure of the source text (ST), which led them to render the sentence ‘The panel discussion was chaired by the President of the Human Rights Council’ into *تراست نقاش اللجنة من قبل رئيس مجلس حقوق الإنسان*. Their attempt to uphold the passive voice structure in translation, although common in English, is less typical in Arabic.

In example 13, participants initially struggled to translate the collocation “adverse impact” from the source text (ST) into an equivalent collocation in Arabic. They rendered it as *التأثير المضاد*, whereas the

more appropriate collocation equivalent is التأثير السلبي. In example 14, participants faced difficulty in maintaining the accurate order of recursive words. Similarly, in example 15, the phrase “people in vulnerable situation” was translated as عل الناس في وضع ضعيف. This translation lacks proper collocation, highlighting the need for a deeper understanding of translation beyond word-for-word rendering. Such understanding requires extensive reading in the field of translation.

Likewise, in example 7, participants translated the word “enjoyment” as الأستمتاع. However, a more suitable collocation-wise translation would be التمتع. Both words are correct, but the latter is a better fit in terms of collocational usage. In example 19, participants adhered to the normal syntactic pattern of Arabic, foregrounding the main verb. However, this did not align with the syntactic pattern of English, resulting in a translation that lacks syntactic correctness. Similarly, in example 19, the phrase “lessons learned” was translated as الدروس المتعلمة, which is accurate but not collocational. In Arabic, the more commonly used phrase is الدروس المستفادة.

Furthermore, an issue regarding accuracy of translation was prevalent in the participants’ work before their exposure to the corpora. For instance, in example 9, participants initially translated the expression “taking note of” into يلاحظ, a common equivalent in English. However, they failed to convey the denotative meaning accurately, as the expression is often translated in UN documents as تحيط علما. A similar situation occurred in example 10, where participants translated “recalling” as يستدعي, while the more accurate translation is إذ تشير إلى.

In example 11, participants translated “panel discussion” as نقاش الخبراء or نقاش اللجنة, assuming that discussion is modified by panel. However, this translation is incorrect. According to the corpus, the accurate translation is حلقة النقاش. Lastly, in example 20, participants rendered parts of the ST literally, resulting in an inaccurate rendition. They translated “International Sign interpretation and real-time captioning” literally as الترجمة الفورية للعلامات الدولية والشرح في الوقت الفعلي. The correct translation is الترجمة الفورية بلغة الإشارة و الترجمة النصية المباشرة.

In conclusion, the use of corpora proved highly beneficial for the participants. It helped them maintain idiomatic translations, utilize proper collocations, and accurately convey intended meanings. These aspects significantly enhanced the quality of their translations, highlighting the value of using specialized corpora in translation practices.

Idiomatic Usage and Collocational Precision:

Example 15 illustrates the importance of idiomatic translation, as the initial translation failed to capture the intended meaning accurately. Exposure to the corpora facilitated a more contextually fitting and collocationally appropriate rendition.

Accuracy of Translation:

Examples 9 and 10 underscore the significance of precise translation, as initial attempts did not accurately convey the denotative meaning. The corpora unveiled more contextually appropriate translations, ensuring accuracy.

Specialized Terminology:

Example 20 showcases how corpora guide the translation of specialized terms, ensuring they are accurately rendered in the target language while avoiding overly literal translations.

In conclusion, the application of corpora significantly enhanced translation quality by addressing key issues related to idiomaticity, collocational accuracy, syntactic patterns, and the precision of translation. By utilizing specialized corpora, novice translators were able to align their translations more closely with native-like expressions and maintain the intended meaning and context.

It's evident from the examples you've provided that using corpora has proven to be highly beneficial for addressing various translation challenges. These benefits can be summarized as follows:

Quantitative Results

Descriptive Statistics

To quantitatively measure the impact of using the corpora, the two translation tasks were marked, and SPSS was used to run paired-samples t-test. To ensure the validity of using the t-test, descriptive statistics were implemented. This includes measuring the frequency, skewness and kurtosis to ensure that the data are normally distributed. Table 2 shows the kurtosis and skewness values.

The information provided in Table 2 indicates that the skewness and kurtosis values for the scores obtained before the participants were exposed to the corpora were measured at .866 and .865, respectively. These values suggest that the data distribution is skewed to the left, indicating a departure from a perfectly normal distribution. However, despite this skewness, the data is still considered to be approximately normally distributed.

Similarly, for the scores acquired after the participants were exposed to the corpora, the skewness and kurtosis values were calculated as .716 and .780. These values also indicate a left skew in the data distribution, signifying that the distribution is not perfectly normal. However, similar to the previous case, the data is still considered to be approximately normally distributed.

Table 2 *Normal distribution: Skewness and Kurtosis*

N	Statistics	
	Beforecorpora	Aftercorpora
Valid	30	30
Missing	0	0
Skewness	.866	.716
Std. error of skewness	.427	.427
Kurtosis	.865	.780
Std. error of kurtosis	.833	.833

Table 3 *Mean and standard deviation*

	Descriptive Statistics				
	N	Minimum	Maximum	Mean	SD
Beforecorpora	30	53.00	85.00	63.9000	7.60376
Aftercorpora	30	60.00	95.00	71.1333	8.35313
Valid N (listwise)	30				

Table 4 *Paired samples statistics*

		Paired Samples Statistics			
		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	beforecorpora	63.9000	30	7.60376	1.38825
	aftercorpora	71.1333	30	8.35313	1.52507

Table 5 *Paired Samples Test*

		Paired Samples Test							
		Paired Differences							
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference		T	df	Sig. (2-tailed)
					Lower	Upper			
Pair 1	beforecorpora – aftercorpora	–7.23333	2.75034	.50214	–8.26033	–6.20634	–14.405	29	.000

Inferential Statistics

A Paired samples t-test was run to see the significance of difference among the participants after having been allowed to use the corpora.

As seen in Table 4 and Table 5, on average the students performed better in the task after using the corpora ($M = 71.31$, $SD = 8.35$) as compared with the task before using the corpora ($M = 64$, $SD = 7.60$). Based on the results of paired samples t-test, $t(29) = 14.41$, $p = .000$, 95% CI $[-8.26, -6.21]$, since the significant value was smaller than alpha, the null hypothesis was rejected. It can be concluded that the treatment (i.e., use of corpora) had a significant effect on the students' achievement.

To sum up, the qualitative and quantitative results showed the significance of using the corpora to teach UN texts.

Conclusion

This study investigated the impact of utilizing the UN corpora for translating UN documents. Both qualitative and quantitative analyses were conducted on responses from 30 students who performed two translation tasks. The findings revealed the significant value of using corpora in UN document translation classes. Corpora serve as authentic source materials, aiding students in comprehending the intricate terminology found in such documents. They also empower students to recognize language usage patterns encompassing syntax, collocations, and discourse structures. This knowledge equips translators to produce translations that accurately convey intended meanings and maintain the appropriate register. In essence, integrating corpora into translation classes enables students to develop essential skills and knowledge to become proficient translators of UN texts. While the use of corpora has benefits, a challenge lies in time consumption during searches. To mitigate this issue, CAT tools offer a solution by expediting and partially automating the process.

References

- Alhassan, A. (2015). A corpora-driven approach for the Sudanese EFL translation classroom: Moving beyond bilingual dictionaries and intuition. *Translation Studies*, 8(2), 135–150.

- Alhassan, A., Sabtan, Y., & Ismail Omar, L. (2021). Using parallel corpora in the translation classroom: Moving towards a corpus-driven pedagogy for Omani translation major students. *Arab World English Journal (AWEJ)*, 12, 40–58. <https://doi.org/10.24093/awej/vol12no1.4>
- Alotaibi, H. M. (2016). AEPC: Designing an Arabic/English parallel corpus. *Research in Corpus Linguistics*, 1–7.
- Alotaibi, H. M. (2017). Arabic-English parallel corpus: A new resource for translation training and language teaching. *Arab World English Journal (AWEJ)*, 8, 73–84.
- Altenberg, B. (1998). On the phraseology of spoken English: The evidence of recurrent word-combinations. *Computational Linguistics*, 24(2), 259–262.
- Baker, M. (1995). Corpora in translation studies: An overview and some suggestions for future research. *Target*, 7(2), 223–243.
- Beeby, A., Sánchez-Gijón, P., & Rodríguez Inés, P. (2009). Corpus use and translating. *Corpus Use and Translating*, 1–164.
- Blanco, M. P. (2009). Translating stance adverbials from English into Spanish: A corpus-based study. *International Journal of Translation*, 21(1–2), 41–56.
- Brown, P. F., Della Pietra, S. A., Della Pietra, V. J., & Mercer, R. L. (1991, June). Word-sense disambiguation using statistical methods. In 29th Annual meeting of the Association for Computational Linguistics (pp. 264–270).
- Doval, I., & Nieto, M. T. S. (Eds.). (2019). Parallel corpora for contrastive and translation studies: New resources and applications (Vol. 90). John Benjamins Publishing Company.
- Fantinuoli, C., & Zanettin, F. (2015). Creating and using multilingual corpora in translation studies. *New directions in corpus-based translation studies*, 1.
- Frankenberg-Garcia, A. (2016). The use of corpus examples for language comprehension and production. *ReCALL*, 26(2), 128–146.
- Hu, K. (2016). *Introducing corpus-based translation studies*. Springer.
- Hu, K., & Zou, S. (2009). Shashibiya xiju yinghan pingxing yuliaoku de chuangujian yuyingyong [The compilation and use of the English-Chinese Parallel Corpus of Shakespeare's Plays]. *Foreign Language Studies*, 5, 64–71.
- Johansson, S. (2007). Using corpora: From learning to research. In *Corpora in the foreign language classroom* (pp. 17–28). Brill.
- Johansson, S., & Hofland, K. (2000). The English-Norwegian parallel corpus: Current work and new directions. In *Multilingual corpora in teaching and research* (pp. 134–147). Brill.
- Kay, M., & Roscheisen, M. (1993). Text-translation alignment. *Computational linguistics*, 19(1), 121–142.
- Kennedy, G. (2014). *An introduction to corpus linguistics*. Routledge.
- Laviosa, S. (1996). *The English Comparable Corpus (ECC): A resource and a methodology for the empirical study of translation*. Ph.D. dissertation, University of Manchester.
- Løken, B. (1997). Expressing possibility in English and Norwegian. *ICAME Journal*, 21, 43–60.
- Sabtan, Y. (2016). Bilingual lexicon extraction from Arabic-English parallel corpora with a view to machine translation. *Arab World English Journal (AWEJ) Special Issue on Translation*, 7–18.
- Qin, H., & Wang, K. (2004). Jiyu Yuliaoku De Fanyi Yuyan Fenxi: Yi 'so...that' De Hanyu Duiying Jiegou Weili [Parallel Corpora-Based Analysis of Translationese: The English 'So...That' Structure and Its Chinese Equivalents in Focus]. *Modern Foreign Languages*, 27(1), 40–48.
- Xiao, R., & Dai, G. (2014). Lexical and grammatical properties of translational Chinese: Translation universal hypotheses reevaluated from the Chinese perspective. *Corpus Linguistics and Linguistic Theory*, 10(1), 11–55.
- Zampieri, M., & Vela, M. (2014, April). Quantifying the influence of MT output in the translators' performance: A case study in technical translation. In *Proceedings of the EACL 2014 Workshop on Humans and Computer-assisted Translation* (pp. 93–98).
- Zanettin, F., Bernardini, S., & Stewart, D. (Eds.). (2003). *Corpora in translator education*. St. Jerome Publishing.

UN Translation Exam Marking Rubric

Criteria	Excellent (90–100)	Good (80–89)	Satisfactory (70–79)	Needs Improvement (50–69)	Inadequate (0–49)
Accuracy	The translation is highly accurate, demonstrating an excellent understanding of the source text with minimal errors. Terminology is precise, and the translation fully conveys the intended meaning.	The translation is mostly accurate, showing a strong comprehension of the source text with few minor errors. There may be occasional terminology discrepancies, but they do not significantly impact understanding.	The translation has some accuracy issues, including noticeable errors in terminology. While the overall meaning is clear, there are areas requiring improvement.	The translation contains significant errors that impact comprehension and terminology precision. The overall meaning is affected.	The translation is largely inaccurate, making it difficult to understand. Key terminology is misused, leading to confusion.
Clarity and Cohesion	The translation is exceptionally clear, well-organized, and maintains cohesion throughout. It adheres to UN style guidelines and mirrors the structure of the source text effectively.	The translation is generally clear and organized but may have minor issues with cohesion. It mostly adheres to UN style guidelines and preserves the structure of the source text.	The translation is somewhat clear, but there are noticeable problems with organization and cohesion. Some deviations from UN style guidelines are present.	The translation lacks clarity and has significant issues with organization and cohesion. It deviates noticeably from UN style guidelines.	The translation is confusing and lacks coherence. It strays significantly from UN style guidelines.
Terminology Precision	The translated text uses precise and appropriate UN terminology consistently, enhancing the overall quality of the translation. Terminology choices align closely with established UN terminology.	The translation employs suitable UN terminology, with minor instances of imprecise or repetitive language. Terminology choices mostly align with established UN terms.	The translation includes basic UN terminology but may lack variety and precision in certain areas. Some terminology choices may deviate from established UN terms.	The translation struggles with appropriate UN terminology, leading to confusion. Terminology choices significantly deviate from established UN terms.	The choice of terminology in the translation is inadequate and significantly impacts comprehension. Frequent deviations from UN terms are observed.
Language Proficiency	The translated text demonstrates a high level of language proficiency, exhibiting mastery in both source and target languages. The language used is clear, precise, and aligns with diplomatic language standards.	The translation reflects a good command of both languages, with minor linguistic issues that do not significantly impede understanding. The language used mostly aligns with diplomatic language standards.	The translation shows adequate language proficiency, but there are noticeable errors and awkward phrasing. Some instances of language use may not fully align with diplomatic language standards.	The translation exhibits limited language proficiency, making it challenging to understand. Frequent deviations from diplomatic language standards are observed.	The language proficiency in the translation is insufficient, hindering comprehension. The language use significantly deviates from diplomatic language standards.