Evaluating Parallel Corpora Based Machine Translation Performance from English into Arabic

تقييم أداء الترجمة الآلية المعتمدة على المتوازيات اللغوية من الإنجليزية إلى العربية

Évaluation des performances de la traduction automatique basée sur des corpus parallèles de l'anglais vers l'arabe

Oussama Chetioui - Laboratory of Translation and Text, Typology - Oran 1

Evaluating Parallel Corpora Based Machine Translation Performance from English into Arabic

تقييم أداء الترجمة الآلية المعتمدة على المتوازيات اللغوية من الإنجليزية إلى العربية

Évaluation des performances de la traduction automatique basée sur des corpus parallèles de l'anglais vers l'arabe

OUSSAMA CHETIOUI
LABORATORY OF TRANSLATION AND TEXT, TYPOLOGY
UNIVERSITY OF ORAN 1

Introduction

Translation is, as it were, the process of recreating a text to parallelize its source, structurally, semantically, contextually, and at many other levels in another language. In machine translation (MT), the technicality of text recreation gains further peculiarity, as the only tool used here is exclusively the machine, and human contribution is either absent or preceding the process itself. In this context, machine translation has received remarkable advances over recent decades. From the early word-for-word correspondence to current approaches that care more about context, MT has reached exciting levels in terms of the processing itself or its output qualities that, in many ways, resemble human translation, and even occasionally outperform it.

Such an evolution should not depart from the keen contribution of parallel corpora approaches adopted in the overall advancement of MT. The sentential alignment approach provides clues for MT to better explore the different languages' ways of meaning expression at the syntax level. This sentence-to-sentence (source-target) approach revolutionized the situation of several MT performances of many languages that suffered clumsiness and ambiguities in earlier approaches. In this vein, Arabic was relatively disregarded when it comes to quality outputs of MT performances. However, since the arrival of what is called in this study "parallel corpora-based machine translation" (henceforth PCBMT), the Arabic renditions of MT witnessed clear refinements and hence serious research was induced to evaluate the differences. However, despite refinements, issues of inconsistencies in meeting the appropriate norms dictated by the Arabic peculiarities are still raised in sentence structure and meaning expression techniques throughout the sentence.

Therefore, there has been a serious tendency to provide PCBMT systems with quantities of parallel corpora and make their online availability easier to recognize. Notwithstanding, the Arabic renditions of PCBMT remained

suffering inconsistency and even translation error consequences in several situations. This study justifies this by the issue of quality, which these inserted parallel corpora bear at levels of structure, stylistics, and context invariance. As a result, it would not be surprising if the ultimate output of PCBMT performance lacks quality. Using an English-into-Arabic translation directionality, this study explores and evaluates the PCBMT performance quality in two different situations. The first situation deals with a virgin text in English, which Arabic translation has not been provided, to evaluate the performances provided by two different PCBMT systems (Neural MT and Statistical MT) in Arabic, and to pinpoint inconsistencies in the translational and the linguistic level. The second situation, however, deals with another English text (of the same context) for which Arabic translation is available online, and which both PCBMT systems tend to use in their processing. The results are then evaluated and analyzed using a range of quantitative tools to gauge the quality of each performance amid both situations.

Using the aforementioned, this study is threefold. First, it embarks on providing descriptive views on the issue of the quality of English-Arabic parallel corpora available for use on the net. Second, it evaluates each PCBMT performance of the corpora used all along the evaluation. Third, after the evaluation results are extracted, this study quantitatively and qualitatively analyzes the obtained results to conclude with an overview of the reality of the PCBMT performance from English into Arabic.

1. Advances in English-Arabic Machine Translation

1.1 Parallelizing Machine Translation

Amid fierce criticism for its early performances, it was an inevitable necessity for machine translation to provide solutions for its clumsy renditions of several key languages at many levels (including lexis, syntax, semantics, and terminology). Therefore, researchers and software developers sought a range of approaches to help machine translation algorithms better recognize the source text's characteristics (linguistically and semantically) and hence provide refined processing throughout the translation process. These approaches gradually evolved from rule-based machine translation to the most recent tendency: context-based machine translation. In the middle was one of the pioneering approaches in the overall study of machine translation, both in theory and practice. Corpus-based machine translation (CBMT) occupied (and occupies) a central position in the overall evolution of machine translation performances. This approach relies on providing machine

translation software with numerous texts (a large corpus) in a plethora of different languages to help them process the source text. This process relies on the existing corpus of the same language in use (Spyns & Odijk, 2012).

Shortly, this corpus-based machine translation approach received crucial strengthening through a parallelizing approach of corpora. Parallel corpora did not merely provide a different machine translation technique, but they moved the overall automated domain of translation to the next level (Dorr, 2011). This highly data-driven approach enables machine translation algorithms to run comparative processing of languages in charge (machine translation directionality) and to search for similar structures and contexts in the already existing parallel corpus. The leading machine translation types of these approaches are Statistical Machine Translation (SMT) and its evolution : Neural Machine Translation (NMT). Since context-based machine translation is still a brand-new approach, particularly in practice, relatively all current machine translation software and online services are either statistical, neural or even a hybrid of both. SMT and NMT operate on the principle of auto-learning, i.e. both approaches embark from the parallel corpus data stored in their systems, and the more data they have, the better the translation quality is. As Forcada (2017) argues in this regard, "in both NMT and SMT, a target sentence is a translation of a source sentence with a certain probability of likelihood" (p. 300).

However, since these approaches rely on a quantitative probabilistic method, their translation output has not always been stable. Occasionally, PCBMT approaches may fail either to transfer the original semantics or to adhere to the target language structure, or at both levels. Such failure can be realized in minoritized languages (where online corpus is scarce) and even in languages with highly different systems (mainly from English) and of less online interest among its own researchers (Arabic is a clear example).

1.2 English-Arabic Parallel Corpora : Alignment Method and Quality Issue

Several alignment methods of English-Arabic parallel corpora are still using word-for-word correspondences, even if the text is segmented into separate sentences (see figure 1 below). Figure 1. Word-for-word Correspondence of an Online Arabic-English Parallel Corpus

Figure 1. Word-for-word Correspondence of an Online Arabic-English Parallel Corpus

Source: Opus Website : TED2013 Resource: https://opus.nlpl.eu/sample/ar&en/TED2013&v1.1/sample

In Figure 1 above, not only is the translation quality lacking in some parts, but a contradiction is also present in the elaboration of this parallel corpus. As shown in Figure 1, the source language is indicated as Arabic, and the target language like English. This is completely erroneous and misleading, as no indicator of a human production of the Arabic language is present. The stylistic semantics are wrong in some parts, and the sentence structuring is clearly not humanly produced. This reiterates the fact that the quality of parallel corpora, particularly in such translation directionality (English-Arabic), should be brought into question and subjected to further evaluation since this parallel corpus is certainly browsed by machine translation software during processing, and hence the output and overall performance will indeed be impacted.

This tendency likely raises the probability of distorting the intended meaning of the whole text and hence impairs the processing of this kind of parallel corpus by any machine translation system. It is noteworthy that (free and online) sources for sufficient quality of English-Arabic parallel corpora are limited. Arguably, the United Nations online system for corpus parallelism is the only guaranteed source that provides good quality English-Arabic parallel corpora, being free of charge for both PCBMTs and other uses. These corpora enjoy good quality because they are humanly translated from English into Arabic (and into all the remaining four languages of the UN). However, these parallel corpora are again limited to the outputs of the organization's proceedings and annual reports, which is insufficient if an evolution of Arabic SMT and NMT is sought.

Since this article focuses attention on a particular machine translation directionality (English into Arabic), it is noteworthy that English-Arabic parallel corpora available on the net for machine translation systems are not always of sufficient quality. Some English-Arabic parallel corpora are merely machine translations of the English text into Arabic, examples of which include highly specialized contexts, such as technical texts. Furthermore, some English-Arabic (and vice versa) parallel corpora suffer from incorrect alignment of sentences. That is to say, target sentences do not align with source sentences (lexically, semantically, terminologically, etc.). Target sentences here are completely other translations of other parts of the text, or even sentences with no existing parallels in the whole text (see Figure 2 below) :

Figure 2. Sentence Misalignment of an English-Arabic Parallel Corpus

```
(src)="4: الدين
(trg)="9: Later,
(src)="5: صنع القرار والتأثير
(trg)="12: The city plan dates back to the 1650s,
(src)="6: معلومات أساسية
(trg)="15: Kokkola is a city of culture that offers a lot to see and experience.
(src)="7: تقع كوكولا في منطقة بوهينما الوسطى ، على ساحل خليج بوتنية .
(trg)="17: The city port is located in the district of Ykspihlaja, and it is one of the busiest
cargo ports in Finland.
(src)="8: أُسست مدينة كوكولا عام 1620 .
(trg)="19: Chemical industry and applied industries are in a central position.
(src)="9: يوجد بالقرب من مركز المدينة التجاري أحد أكبر تشكيلات البيوت الخشبية التاريخية .
(trg)="25: The rising of the land has been a key factor in the history of Kokkola.
(src)="10: يعود تاريخ مخطط المدينة إلى العام 1650 وهو يغطي 12 حياً فيها مئات البيوت الخشبية ومباني الباحات ، ويعود أقدمها إلى القرن 1600 .
(trg)="26: It has shaped the environment significantly and affected the development of the port
and commerce, for example.
(src)="11: وهو مدينة ثقافة ويوجد فيها كثير من المعالم والتجارب .
(trg)="28: The tar export business, later to become crucial to the history of Kokkola, began as
early as in the 16th century.
(src)="12: تبلغ المسافة من كوكولا إلى هلسنكي 482 كم وإلى يوفسكيلا 242 كم وإلى أولو 198 كم وإلى بيترساري 36 كم وإلى سينايوكي 141 كم وإلى تامبري 302 كم وإلى فازا 121 كم .
(trg)="29: On 7th September 1620, the Swedish king Gustav Adolf II signed a document
transforming the small Ristiranta village, known for fishing and farming, into a town.
(src)="13: وهو أحد أكثر موانئ فنلندا حركة في نقل البضائع Ykspihlaja يقع ميناء البحري في حي اكسبهلايا .
(trg)="32: There were shipyards in Kaustarinlahti, Mustakari and Soldatskär, for example.
(src)="14: يتألف أساس الحياة الاقتصادية في كوكولا من الصناعة الدولية الكبرى التي تحتل المركز الرئيسي فيها الصناعة الكيميائية ومشتقاتها التطبيقية .
(trg)="38: The rapid economic development stopped in the middle of the 19th century, only to
bloom again in the late 1800s, spurred on by industrialisation.
(src)="15: كذلك تعتبر كوكولا مدينة تجارة مهمة .
(trg)="45: The railway station is located in the city centre.
(src)="18: التاريخ
(trg)="62: Religion
```

Source: Opus Website: infopankki Resource: https://opus.nlpl.eu/sample/ar&en/infopankki&v1/sample

As Figure 2 shows, besides the low translation quality, a large part of sentences is aligned improperly. This is utterly confusing to the machine translation systems using this parallel corpus, as not only is the translation quality a problem here, but an issue of non-translation appears in such a situation. PCBMT systems' performance would be highly impacted, particularly since once again the misleading translation directionality is also present in this parallel corpus.

1.3. Systemic Approach to Evaluating PCBMT

A variety of studies have already addressed machine translation evaluation methods and metrics. At one time, it was argued that more has been written on machine translation evaluation than on machine translation itself (Hovy, 2002). However, Hovy argues that one machine translation evaluation human metric stood for a long time and received general agreement among machine translation evaluators. This particular metric is built upon two major scales: fluency and fidelity. Fluency, as it may indicate, refers to the machine translation output's intact expression of meaning flow through the sentence without incongruities. Fidelity, on the other hand, refers to the MT output's success in transferring all contextual nuances intended in the original text.

A well-translated text reads properly to its ultimate users (Frankenberg-Garcia, 2021). Therefore, it is of paramount importance for the English-to-Arabic machine translation directionality to consider all necessary components of both the input's and output's semantics and structure. Hence, the text sentencing in PCBMT systems should produce useful outcomes, not the opposite, for this translation directionality. Given that the text's context relies on its sentences' coherence and cohesion to function as communicative material, Arabic-translated sentences, whether automated or original, need to meet at least the minimum quality levels of both semantics (fidelity) and language structure (fluency). It is not uncommon for SMT and NMT systems to fail to produce a coherent full text in Arabic. The primary reason behind this failure could be a lack of similar parallel corpora or, more often, low-quality parallel corpora stored in their systems. This is exacerbated by the fact that many SMT and NMT systems adopt an example-based method in their processing of English-Arabic directionality, increasing the likelihood of an erroneous rendition.

Using the aforementioned considerations, this article starts from the premise that the quality of online English-Arabic parallel corpora is not always adequate, to provide a comprehensive evaluation of PCBMT translation systems' performance. The idea revolves around testing the

performance of two major online (free) machine translation services (Google Translate and Reverso Translation) from English into Arabic amid the existence and non-existence of the input's parallel text. Analysis of each performance outcome is also provided in different forms (text, tables, charts, figures). The ultimate objective is to provide a more systemic approach to evaluating PCBMT performance in the English-Arabic directionality and to assist further research (which is required) to use the collected data and results.

2. Methodology, corpus and procedure

This section describes the study corpora and the method used to compare the performances of each translation service used in the study. It also analyzes the PCBMT performance variations when translating a source text that already has an online-available parallel text (situation 2) and another source that does not (situation 1).

This study operates on two main levels. First, a source text (an excerpt) provided by the African Union report on gender equality published in February 2020 (henceforth, Corpus 1), whose original (English) text has not been subject to any sort of translation, human or machine, into Arabic. Second, it uses an excerpt from a United Nations online report published on January 4, 2021 (henceforth, Corpus 2) in the same context of gender equality available on the organization's website (web corpora), and its human translation into Arabic. Despite having the United Nations' web corpora mostly available in all six languages adopted by the organization, this study's analysis focuses on the English-Arabic corpus exclusively. The choice of both corpora adopted in this study was random, as the main objective is to evaluate PCBMT performance that is universal regardless of corpus particularization.

Given that the UN is a professional organization that seeks to meet all requirements of the six languages adopted, it is then assumed that machine translation (MT) did not intervene at any part of the humanly translated texts.

The volume of the corpora selected is based on the two online translation services' allowed number of characters for a translation to be fully performed. Although Google Translate allows up to 5000 characters, Reverso Translation only allows 2000 characters. As a result, it was necessary to limit the source text to meet both translation services' requirements for a fair comparison and for more reliable results to use in analysis.

The choice of Google Translate and Reverso Translation was based on three reasons. Firstly, both MT systems are free of charge and widely used by the public, particularly for English-to-Arabic translations. Secondly, this

choice allows this study to compare two different PCBMT systems: Google Translate with its purely neural approach and Reverso Translation with its hybridization of SMT and rule-based performance. Thirdly, both translation services rely upon the UN outputs (parallel corpora) in their processing. Note that this study does not intend to develop any of these systems; it only uses them as a tool and platform where their English-Arabic translation performances can be extracted and evaluated (in both situations: existence and non-existence of the parallel text).

After obtaining the MT outputs of Corpus 1 and Corpus 2, they are segmented into sentences, so that each sentence is evaluated separately. Then, an evaluation is held using the following key:

- (F): Translated sentence meets Fluency
- (Fd): Translated sentence meets Fidelity
- (√): Correct translation √
- (x): Erroneous translation

The evaluation is conducted by the author (humanly), segmenting the text into sentences (source aligning with target). If a translated sentence meets fluency (F) but fails to meet fidelity (Fd), it is evaluated as a Contextual Error. If a translated sentence meets fidelity (Fd) but fails to meet fluency (F), it is evaluated as a Structural Error. If a translated sentence fails to meet either fluency (F) or fidelity (Fd), or expresses a clearly contradictory meaning to the original, it is evaluated as an Erroneous Translation. If a translated sentence meets both fluency (F) and fidelity (Fd), it is evaluated as a Correct Translation.

Figure 3 Material used for the study
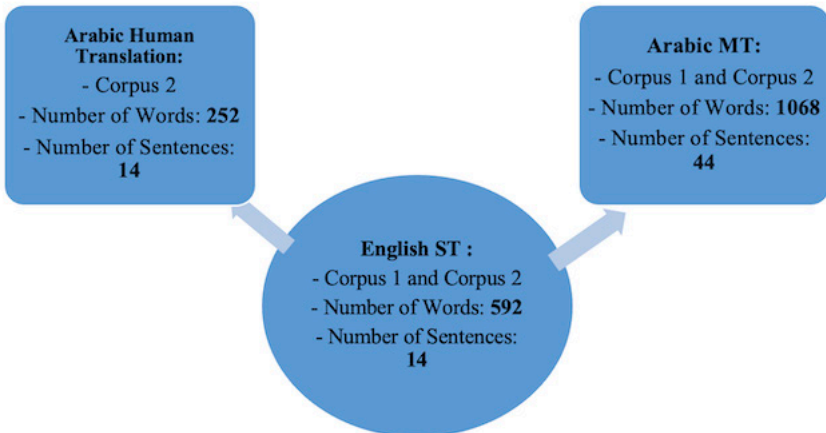
3. Findings and Analysis

3.1 PCBMT Performance Evaluation (no parallel text used)

Corpus 1 was inserted for MT in Google Translate and Reverso Translation consecutively. Then, an evaluation of each translated sentence performed by both MT systems was conducted using the procedure defined. Evaluation findings of each performance are categorized as errors in fluency, fidelity, and translation error. Furthermore, each translated sentence is evaluated for structural and contextual errors in comparison. Error scores are presented in different forms, such as tables and charts. Further analysis is also provided at other levels, evaluating both NMT and SMT performances. Figure 4 below provides details on Corpus 1 (AU report).

Figure 4. Corpus 1 details used in the study
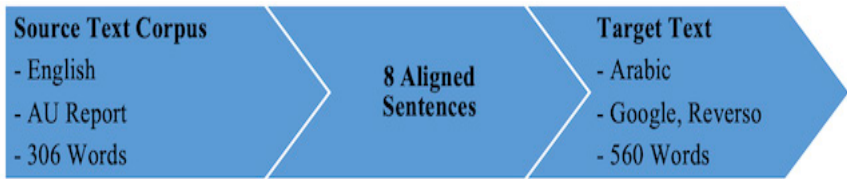
Figure 5 below shows the original text in English used in the evaluation (corpus 1)

Figure 5. Source text in English (Corpus 1)

IV.   Gender Equality and Women's Empowerment

a.   Advancing Gender Agenda

255.   The Commission presented the 4th Edition of the AU Gender Scorecard, which were presented to 17 Heads of State and Government at the 32nd Ordinary Session of the Assembly, held in February 2019, in Addis Ababa, in recognition of progress made in implementing Gender Equality and Women's Empowerment (GEWE) Commitments related to the AU theme of the year. In 2019, the Commission continued to promote the integration of the AU Common Africa Position on GEWE in international platforms and participated in the 63rd Session of the United Nations Commission on the Status of Women (CSW63) as well as the 74th session of the UN General Assembly held at the UN Headquarters in March and September 2019, respectively.

256.   The 10th Anniversary of the AU Gender Pre-Summit was celebrated in an event co-hosted with the Egypt National Council for Women February, 2019, in Cairo, Egypt. In the same vein, in July, 2019, a Pre-Summit on "*Advancing Women and Girls Empowerment*" was convened in collaboration with the Ministry of Women Empowerment and Child Protection of the Republic of Niger together with the AU International Centre for Girls and Women's Education in Africa (AU/ CIEFFA) in Niamey, Niger, on the margins of the First Mid-Year Coordination Meeting of the AU and the Regional Economic Communities (RECs). A Solidarity Mission in Sudan was undertaken in collaboration with the Office of the Special Envoy on Women, Peace and Security to support the women of Sudan in their endeavours for a democratic, peaceful and inclusive Sudan that incorporates women in decision-making at all levels. An in-country advocacy mission as well as a training on the ratification, domestication and implementation of the Maputo Protocol followed in November, 2019, in Khartoum, Sudan to support the institutionalisation of women's rights in the new government of Sudan.

Source: AU Report: African Union, 2020, p 57

In table 1 below, the PCBMT performance of Google Translate of corpus 1 into Arabic is evaluated:

Table 1. Evaluating Google Translate Performance amid the absence of the input's parallel text in Arabic

Source Text	Human Translation	Machine Translation
IV. Gender Equality and Women's Empowerment	رابعا - المساواة بين الجنسين وتمكين المرأة. (√)	رابعا - المساواة بين الجنسين وتمكين المرأة. (√)
a. Advancing Gender Agenda	ألف - النهوض بجدول الأعمال الجنساني. (√)	أ. النهوض بجدول الأعمال الجنساني. (F)
The Commission presented the 4th Edition of the AU Gender Scorecard, which was presented to 17 Heads of State and Government at the 32nd Ordinary Session of the Assembly, held in February 2019, in Addis Ababa, in recognition of progress made in implementing Gender Equality and Women's Empowerment (GEWE) Commitments related to the AU theme of the year.	قدمت المفوضية الإصدار الرابع من بطاقة قياس الأداء الجنسانية للاتحاد الأفريقي، والتي تم تقديمها إلى 17 من رؤساء الدول والحكومات في الدورة العادية الثانية والثلاثين للمؤتمر، التي عقدت في فبراير 2019، في أديس أبابا، تقديرا للتقدم المحرز في تنفيذ المساواة بين الجنسين وتمكين المرأة (GEWE) الالتزامات المتعلقة بموضوع الاتحاد الأفريقي لهذا العام. (Fd)	قدمت المفوضية الطبعة الرابعة من سجل النتائج الجنسانية للاتحاد الأفريقي، التي قدمت إلى رؤساء الدول والحكومات 17 الدورة العادية الثانية والثلاثين للجمعية، المعقودة في شباط/فبراير 2019، في أديس أبابا، اعترافا بالتقدم المحرز في تنفيذ التزامات المساواة (GEWE) بين الجنسين وتمكين المرأة المتصلة بموضوع الاتحاد الأفريقي لهذا العام. (Fd)
In 2019, the Commission continued to promote the integration of the AU Common Africa Position on GEWE in international platforms and participated in the 63rd Session of the United Nations Commission on the Status of Women (CSW63) as well as the 74th session of the UN General Assembly held at the UN Headquarters in March and September 2019, respectively.	في 2019، واصلت المفوضية تعزيز تكامل الموقف الأفريقي الموحد في GEWE للاتحاد الأفريقي بشأن المحافل الدولية وشاركت في الدورة 63 للجنة الأمم المتحدة المعنية بوضع المرأة (CSW63) وكذلك الدورة 74 للأمم المتحدة العامة التي عقدت في مقر الأمم المتحدة في آذار/مارس وأيلول/سبتمبر 2019 على التوالي. (Fd)	وفي 2019، واصلت اللجنة تعزيز إدماج موقف الاتحاد الأفريقي في البرامج GEWE المشترك بشأن الدولية، وشاركت في الدورة الثالثة والستين للجنة وضع المرأة التابعة وكذلك (CSW63) للأمم المتحدة الدورة الرابعة والسبعين للجمعية العامة للأمم المتحدة التي عقدت في مقر الأمم المتحدة في آذار/مارس وأيلول/سبتمبر 2019 على التوالي. (F)

Source Text	Human Translation	Machine Translation
The 10th Anniversary of the AU Gender Pre-Summit was celebrated in an event co-hosted with the Egypt National Council for Women February 2019, in Cairo, Egypt.	وقد تم الاحتفال بالذكرى السنوية العاشرة لمؤتمر القمة السابق لقضايا الجنسين الذي عقده الاتحاد الأفريقي في مناسبة شارك في استضافتها المجلس الوطني للمرأة في مصر في شباط/فبراير، 2019 . (√)، في القاهرة، مصر	وقد تم الاحتفال بالذكرى السنوية العاشرة للقمة السابقة للقمة الجنسانية للاتحاد الأفريقي في حدث استضافته بالاشتراك مع المجلس القومي للمرأة في مصر في فبراير 2019 (F). في القاهرة، مصر
In the same vein, in July 2019, a Pre-Summit on "Advancing Women and Girls Empowerment" was convened in collaboration with the Ministry of Women Empowerment and Child Protection of the Republic of Niger together with the AU International Centre for Girls and Women's Education in Africa (AU/CIEFFA) in Niamey, Niger, on the margins of the First Mid-Year Coordination Meeting of the AU and the Regional Economic Communities (RECs).	في نفس السياق، في تموز/يوليه، 2019، عقد مؤتمر قمة مسبق بشأن "النهوض بتمكين المرأة والفتاة" بالتعاون مع وزارة تمكين المرأة وحماية الطفل في جمهورية النيجر مع المركز الدولي لتعليم الفتيات والنساء في أفريقيا التابع /للاتحاد الأفريقي (الاتحاد الأفريقي CIEFFA) في نيامي، النيجر، على هامش الاجتماع التنسيقي الأول لمنتصف السنة للاتحاد الأفريقي والجماعات الاقتصادية الإقليمية (REC). (√)	في نفس السياق، في يوليو 2019، تم عقد قمة مسبقة على "النهوض بتمكين المرأة والفتاة" بالتعاون مع وزارة تمكين المرأة وحماية الطفل في جمهورية النيجر بالاشتراك مع المركز الدولي لتعليم الفتيات والنساء في أفريقيا في نيامي، النيجر، على هامش (AU/ CIEFFA) الاجتماع التنسيقي الأول لمنتصف السنة للاتحاد الأفريقي والمجتمعات الاقتصادية الإقليمية (REC). (√)
A Solidarity Mission in Sudan was undertaken in collaboration with the Office of the Special Envoy on Women, Peace and Security to support the women of Sudan in their endeavours for a democratic, peaceful and inclusive Sudan that incorporates women in decision-making at all levels.	وتم الاضطلاع ببعثة تضامن في السودان بالتعاون مع مكتب المبعوث الخاص المعني بالمرأة والسلام والأمن لدعم المرأة السودانية في مساعيها الرامية إلى إقامة سودان ديمقراطي وسلمي وشامل يضم المرأة في عملية صنع القرار على جميع المستويات. (√)	تم الاضطلاع ببعثة تضامن في السودان بالتعاون مع مكتب المبعوث الخاص بشأن المرأة والسلام والأمن لدعم نساء السودان في جهودهن نحو سودان ديمقراطي وسلمي وشامل يضم المرأة في صنع القرار على جميع المستويات. (√)
An in-country advocacy mission as well as a training on the ratification, domestication and implementation of the Maputo Protocol followed in November 2019, in Khartoum, Sudan to support the institutionalisation of women's rights in the new government of Sudan.	وتبع ذلك في تشرين الثاني/نوفمبر 2019، في الخرطوم بالسودان، بعثة للدعوة داخل البلد، فضلا عن تدريب على التصديق على بروتوكول مابوتو وإدماجه في المجتمع المحلي وتنفيذه، وذلك لدعم إضفاء الطابع المؤسسي على حقوق المرأة في حكومة السودان الجديدة. (F)	تبع ذلك بعثة مناصرة داخل الدولة بالإضافة إلى تدريب على التصديق على بروتوكول مابوتو وإدماجه وتنفيذه في نوفمبر 2019 في الخرطوم، السودان لدعم إضفاء الطابع المؤسسي على حقوق المرأة في حكومة السودان الجديدة. (F)

In table 2 below, the PCBMT performance of Reverso Translation of the Corpus 1 into Arabic is evaluated:

Table 2. Evaluating Reverso Translation Performance amid the absence of the input's parallel text in Arabic

Source Text	Translation Output
IV. Gender Equality and Women's Empowerment	رابعا - المساواة بين الجنسين وتمكين المرأة. (√)
b. Advancing Gender Agenda	باء - النهوض بجدول الأعمال الجنساني. (√)
The Commission presented the 4th Edition of the AU Gender Scorecard, which were presented to 17 Heads of State and Government at the 32nd Ordinary Session of the Assembly, held in February 2019, in Addis Ababa, in recognition of progress made in implementing Gender Equality and Women's Empowerment (GEWE) Commitments related to the AU theme of the year.	وقدمت اللجنة الطبعة الرابعة من سجل النتائج الجنسانية للاتحاد الأفريقي، والتي قدمت إلى 17 من رؤساء الدول والحكومات في الدورة العادية الثانية والثلاثين للجمعية، المعقودة في شباط/فبراير 2019، في أديس أبابا، اعترافًا بالتقدم المحرز في تنفيذ التزامات المساواة بين الجنسين وتمكين المرأة (GEWE) المتعلقة بموضوع الاتحاد الأفريقي لهذا العام. (Fd)
In 2019, the Commission continued to promote the integration of the AU Common Africa Position on GEWE in international platforms and participated in the 63rd Session of the United Nations Commission on the Status of Women (CSW63) as well as the 74th session of the UN General Assembly held at the UN Headquarters in March and September 2019, respectively.	وفي عام 2019، واصلت اللجنة تعزيز إدماج موقف الاتحاد الأفريقي المشترك بشأن GEWE في المنصات الدولية، وشاركت في الدورة الثالثة والستين للجنة وضع المرأة التابعة للأمم المتحدة (CSW63) وكذلك الدورة الرابعة والسبعين للجمعية العامة للأمم المتحدة التي عقدت في مقر الأمم المتحدة في آذار/مارس وأيلول/سبتمبر 2019 على التوالي. (F)
The 10th Anniversary of the AU Gender Pre-Summit was celebrated in an event co-hosted with the Egypt National Council for Women in February, 2019, in Cairo, Egypt.	وقد تم الاحتفال بالذكرى السنوية العاشرة لمؤتمر القمة السابق لقضايا الجنسين الذي عقده الاتحاد الأفريقي في مناسبة شارك في استضافتها المجلس الوطني للمرأة في مصر في شباط/فبراير 2019، في القاهرة، مصر. (√)

Source Text	Translation Output
In the same vein, in July, 2019, a Pre-Summit on "Advancing Women and Girls Empowerment" was convened in collaboration with the Ministry of Women Empowerment and Child Protection of the Republic of Niger together with the AU International Centre for Girls and Women's Education in Africa (AU/ CIEFFA) in Niamey, Niger, on the margins of the First Mid-Year Coordination Meeting of the AU and the Regional Economic Communities (RECs).	في نفس السياق، في تموز/يوليه 2019، عقد مؤتمر قمة مسبق بشأن "النهوض بتمكين المرأة والفتاة" بالتعاون مع وزارة تمكين المرأة وحماية الطفل في جمهورية النيجر مع المركز الدولي لتعليم الفتيات والنساء في أفريقيا التابع للاتحاد الأفريقي (الاتحاد الأفريقي/CIEFFA) في نيامي، النيجر، على هامش الاجتماع التنسيقي الأول لمنتصف السنة للاتحاد الأفريقي والجماعات الاقتصادية الإقليمية (REC). (√)
A Solidarity Mission in Sudan was undertaken in collaboration with the Office of the Special Envoy on Women, Peace and Security to support the women of Sudan in their endeavours for a democratic, peaceful and inclusive Sudan that incorporates women in decision-making at all levels.	وتم القيام ببعثة تضامن في السودان بالتعاون مع مكتب المبعوث الخاص المعني بالمرأة والسلام والأمن لدعم المرأة السودانية في مساعيها الرامية إلى إقامة سودان ديمقراطي وسلمي وشامل يضم المرأة في عملية صنع القرار على جميع المستويات. (√)
An in-country advocacy mission as well as a training on the ratification, domestication and implementation of the Maputo Protocol followed in November 2019, in Khartoum, Sudan to support the institutionalisation of women's rights in the new government of Sudan.	وتبع ذلك في تشرين الثاني/نوفمبر 2019، في الخرطوم بالسودان، بعثة للدعوة داخل البلد، فضلا عن تدريب على التصديق على بروتوكول مابوتو وإدماجه وتنفيذه، وذلك لدعم إضفاء الطابع المؤسسي على حقوق المرأة في حكومة السودان الجديدة. (F)

3.2 Analysis

Table 3 below uses translation errors per sentence as a scale to score each performance of the aforementioned translation services:

Table 3. Scoring Google Translate and Reverso Translation Performances of Corpus 1 in Arabic

Translation Service	Fluency Translation Error	Fidelity Translation Error	Correct Translation Score	Translation Error Percentage
Google Translate	3	3	2	37.50 %
Reverso Translation	1	2	5	25%
Total	4	5	7	31.25 %

As can be seen in the two performances of each translation service above, the translation output showed a clear disparity. Google Translate's performance scored the higher translation error percentage (37.50%), providing varying performance with three fluency errors and three fidelity errors. Hence, out of eight sentences, Google Translate succeeded in providing two fully correct translated sentences. The GNMT (Google Neural Machine Translation) approach here showed a swinging performance when it comes to adhering to Arabic structurality and failed to properly structure three out of the eight processed sentences. Some parts of the text remained untranslated and preserved their original language form (as seen in Tables 4 and 5 below). GNMT, in the absence of the input's parallel text in Arabic, failed to provide a corresponding translation for some key terms of the topic. For instance, it translated the term "Gender" into "النوع الاجتماعي", which is confusing and even directs the reader into a different context, particularly since this term appeared in the title. Such an error, along with other errors on both scales, fluency and fidelity, shows how challenging it is for GNMT (and NMT in general) to meet the requirements for context invariance when translating from English into Arabic in the absence of a corresponding parallel text for the source text in use.

Reverso Translation's performance, however, was slightly better. It scored a lower translation error percentage (25%). With one fluency error in eight translated sentences, this PCBMT service proved to perform better in terms of sentence restructuring. Reverso Translation failed twice on the fidelity

scale (2 errors) to meet the original text meaning. These errors were caused by a failure to translate an abbreviation (non-translation), which is directly connected to context invariance, and a mistranslation of a key term in the second error. Succeeding in providing five correct sentences, Reverso proved to perform well in the English-Arabic translation directionality in this particular situation (non-existence of parallel text). The use of a hybrid model of rule-based and statistical approaches largely guided the performance towards some invariance. Despite the aforementioned, Reverso Translation also failed to meet some necessary terminological correspondence, such as "pre-summit" translated into "القمة السابق", which is inconsistent with the intended terminological meaning, besides being a key term opening the sentence and cohering it to the overall context. Using the data collected from Reverso's performance in this translation directionality (English-Arabic) amid the non-existence of an input's parallel text, it is argued that SMT still provides better Arabic performances (compared to NMT) if the source is in English.

Tables 4 and 5 below present a comprehensive analysis of the translation performance provided by Google Translate and Reverso Translation respectively:

Table 4. Comprehensive Analysis of Corpus 1 PCBMT performance provided by Google Translate

Google Translate		
Category	**Source Text**	**MT Output**
Translation Loss	in, of, an, a, that, mission	بين، أعمال، من، التي، عام، على، للقمة، و، ذلك، إضفاء، الطابع.
Translation Gain		بين، أعمال، من، التي، عام، على، للقمة، و، ذلك، إضفاء، الطابع.
Terminological Errors	Gender, assembly, promote, pre-summit, communities, advocacy, domestication	النوع الاجتماعي، مؤتمر، تعزيز، القمة السابقة، مجموعات، مناصرة، إدماجه
Non-translation	GEWE, CSW63, CIEFFA, REC, AU	GEWE, CSW6, CIEFFA, REC, AU

Table 5. Comprehensive Analysis of Corpus 1 PCBMT Performance provided by Reverso Translation

Reverso Translation		
Category	**Source Text**	**MT Output**
Translation Loss	at, that, in, of, to, an, with, together, a, mission.	
Translation Gain		بين، الأعمال، قد، شباط، و، التابعة، التي، أذار، أيلول، على، مؤتمر، في، الذي، عقده، تموز، إقامة، عملية، تشرين الثاني، ذلك، إضفاء، الطابع، المجتمع، المحلي.
Terminological Errors	Commission, platforms, domestication, promote, integration, pre-summit, event, domestication.	اللجنة، برامج، إدماجه، تعزيز، إدماج، مؤتمر القمة السابق، مناسبة، إدماجه في المجتمع المحلي.
Non-translation	GEWE, CSW63, CIEFFA, REC	

3.2.1 Linguistic Appropriateness

3.2.1.1 Structural Consistency

Arabic is a very different language from English. Varieties between both languages' systems have always been an inspiration for research and analysis. For a translation to be referred to as linguistically appropriate, a strict adherence to the target language's prevailing norms is an inevitable necessity. Therefore, PCBMT systems are required to provide an output that conforms to the ultimate users' linguistic norms. However, this is not always the case even in some human translations. In Corpus 1 PCBMT performance evaluation, a range of inconsistencies appeared throughout the evaluation process. Table 6 below inventories these inconsistencies:

Table 6. PCBMT Performance Inconsistencies to Arabic Structural Norms

ST Item	PCBMT System	PCBMT Output	Possible Translation
17 heads of state	GNMT	من رؤساء الدول 17	رئيس دولة 17
Were presented	GNMT/SMT	تم تقديمها	قُدمت
Was convened	GNMT	تم الاحتفال	أُحتفل
Was convened	GNMT	تم عقد	عُقد

ST Item	PCBMT System	PCBMT Output	Possible Translation
Was undertaken	GNMT/SMT	تم الاضطلاع	أُضطلع
In-country	GNMT/SMT	داخل الدولة	وطني/محلي
Held	SMT	المعقودة	المُنعقدة
In recognition of	SMT	اعترافا	عرفانا
Mid-year	SMT	لمنتصف السنة	نصف السنوي
Domestication	SMT	إدماجه في المجتمع المحلي	توطين
Women's Empowerment Commitments	GNMT	تمكين المرأة الالتزامات	إلتزامات تمكين المرأة
Total of inconsistencies		11	

3.2.1.2 Context

Context in Arabic is highly affected by structural invariance. That is to say, the Arabic normative tool for maintaining context invariance is the perfection of structurality. Therefore, any erroneous structural element produced would impact the whole context of the sentence or even the paragraph. Procedurally, both PCBMT performances did well in preserving the text's context invariance. However, some structural errors could lead a non-specialized MT user to think of different contexts at once. As referred to in Table 1 above, the GNMT output of the term "Domestication" does not necessarily indicate its intended meaning in this context. While the tenor was about applying the Maputo Protocol locally, the MT output clearly lacked a supporting charge to contextually correspond to the original well-expressed meaning. Alternatively, the SMT system of Reverso Translation opted to enhance the translation output with two items to attempt to correspond to the intended charge. Another key example can be seen in the translation performance of the sentence "Women's Empowerment Commitments," where the PCBMT output clearly lacked the structural consistency prevalent in Arabic. The output conveys a different charge, transforming the action as if the women have commitments to be empowered to meet. Although it is not well-structured, this is the only contextual charge conveyed if this sentence is read separately, which is clearly contradictory to the original context. This indicates that if a parallel text is absent for the input, and while GNMT may outperform in structurality, SMT systems are more concerned with context invariance than NMT systems would be.

3.3 PCBMT Performance Evaluation (with parallel text used)

Corpus 2 was inserted for MT in Google Translate and Reverso Translation consecutively. Then, an evaluation of each translated sentence performed by both machine translation systems was conducted using the procedure defined in section 3.1. Evaluation findings of each performance are categorized based on errors in fluency, fidelity, and translation error of each performance. Furthermore, each translated sentence is evaluated for structural and contextual errors in comparison to the human translation of Corpus 2. Error scores are classified in tables and charts. Further analysis is provided at other levels, evaluating both NMT and SMT performances in recalling the appropriate translation. Figure 6 below shows the English source text (Corpus 2) and its human translation in Arabic:

Original Text and its Parallel in Arabic (human translation)

Source Text	MT Output
Gender Equality The unfinished business of our time Women and girls represent half of the world's popu-lation and, therefore, also half of its potential. Gender equality, besides being a fundamental human right, is essential to achieve peaceful societies, with full hu-man potential and sustainable development. Moreo-ver, it has been shown that empowering women spurs productivity and economic growth. Unfortunately, there is still a long way to go to achieve full equality of rights and oppor-tunities be-tween men and women, warns UN Women. There-fore, it is of paramount importance to end the multi-ple forms of gender violence and secure equal access to quality education and health, economic re-sources and participation in political life for both women and girls and men and boys. It is also essential to achieve equal opportuni-ties in access to employment and to positions of leadership and decision-making at all le-vels.	المساواة بين الجنسين الأعمال الغير منتهية في عصرنا تمثل النساء والفتيات نصف سكان العالم، وبالتالي نصف إمكانياته. إن المساواة بين الجنسين، إلى جانب كونها حق أساسي من حقوق الإنسان، أمر ضروري لتحقيق السلام في المجتمعات وإطلاق إمكانيات المجتمع الكاملة. وعلاوة على ذلك، فقد ثبت أن تمكين المرأة يحفز الإنتاجية والنمو الاقتصادي. ولسوء الحظ، لا يزال هناك طريق طويل لنقطعه لتحقيق المساواة الكاملة في الحقوق والفرص بين الرجال والنساء، كما تحذر هيئة الأمم المتحدة للمرأة. ولذلك، من الأهمية بمكان إنهاء الأشكال المتعددة للعنف بين الجنسين وضمان المساواة في الحصول على التعليم الجيد والصحة والموارد الاقتصادية والمشاركة في الحياة السياسية لكل من النساء والفتيات والرجال والفتيان. كما أنه من الضروري تحقيق تكافؤ الفرص في الوصول إلى الوظائف والمناصب القيادية وصنع القرار على جميع المستويات

Source Text	MT Output
Guterres highlights that gender equality is more im-portant than ever if we are to create prosperous econ-omies and a healthy planet. However, he admits that we face a critical challenge: an alarming $360 billion annual gender gap by 2030. To revert this trend, he has identified five key areas that need joint action: Investing in women, ending poverty, implementing gender-responsive financing, shifting to a green economy and care society and sup-porting feminist change-makers. The UN Secretary-General, Mr. António Guterres has stated that achieving gen-der equality and empowering women and girls is the unfinished business of our time, and the greatest human rights challenge in our world. The United Nations and women UN support for the rights of women began with the Organization's founding Charter. Among the purpos-es of the UN declared in Article 1 of its Charter is "To achieve international co-operation … in promot-ing and encouraging res-pect for human rights and for fundamen-tal freedoms for all without distinction as to race, sex, language, or religion." Within the UN's first year, the Econo-mic and Social Council established its Commission on the Status of Women, as the principal global policy-making body dedicated exclusively to gender equality and ad-vancement of women. Among its earliest accom-plishments was ensuring gender neutral language in the draft Uni-versal Declaration of Human Rights.	صرح الأمين العام للأمم المتحدة، السيد أنطونيو غوتيريس، أن تحقيق المساواة بين الجنسين وتمكين النساء والفتيات هو عمل غير مكتمل في عصرنا، وأكبر تحدٍ لحقوق الإنسان في عالمنا. الأمم المتحدة والمرأة بدأ دعم الأمم المتحدة لحقوق المرأة مع الإطار الدولي المعلن في ميثاق الأمم المتحدة. ومن بين مقاصد الأمم المتحدة المعلنة في المادة 1 من ميثاق الأمم المتحدة " لتحقيق التعاون الدولي ...على تعزيز احترام حقوق الإنسان والحريات الأساسية للناس جميعاً والتشجيع على ذلك إطلاقاً بلا تمييز بسبب الجنس أو اللغة أو الدين ولا تفريق بين الرجال والنساء". وفي العام الأول للأمم المتحدة، أنشأ المجلس الاقتصادي والاجتماعي لجنة وضع المرأة، بصفتها الهيئة العالمية الرئيسية لصنع السياسات المتعلقة حصرا بتحقيق المساواة بين الجنسين والنهوض بالمرأة. وكان ومن أوائل أنجازاتها هو ضمان لغة محايدة بين الجنسين في مشروع الإعلان العالمي لحقوق الإنسان.

Source: UN Report : United Nations, 2020

Figure 7 below provides quantitative details on corpus 2, and the output provided by Google Translate and Reverso Translation when translating into Arabic, along with the human translation data:

Three-dimensional quantitative data of corpus 2 and its human and PCBMT in Arabic
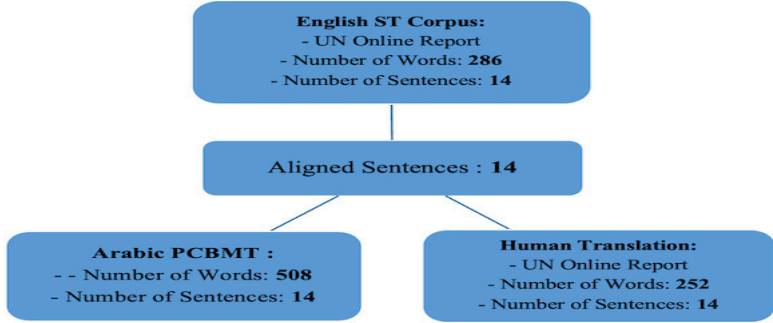
In table 7 below, the PCBMT performance of Google Translate of corpus 2 into Arabic is evaluated:

Table 7. Evaluating Google Translate Performance amid the existence of the input's parallel text in Arabic

Source Text	Translation Output
Gender Equality	المساواة بين الجنسين. (√)
The unfinished business of our time	الأعمال غير المنتهية في عصرنا. (√)
Women and girls represent half of the world's population and, therefore, also half of its potential.	تمثل النساء والفتيات نصف سكان العالم، وبالتالي نصف إمكاناته أيضًا. (√)
Gender equality, besides being a fundamental human right, is essential to achieve peaceful societies, with full human potential and sustainable development.	تعد المساواة بين الجنسين، إلى جانب كونها حقًا أساسيًا من حقوق الإنسان، ضرورية لتحقيق مجتمعات مسالمة تتمتع بإمكانيات بشرية كاملة وتنمية مستدامة. (√)
Moreover, it has been shown that empowering women spurs productivity and economic growth.	علاوة على ذلك، فقد ثبت أن تمكين المرأة يحفز الإنتاجية والنمو الاقتصادي. (√)
Unfortunately, there is still a long way to go to achieve full equality of rights and opportunities between men and women, warns UN Women.	لسوء الحظ، لا يزال هناك طريق طويل لنقطعه لتحقيق المساواة الكاملة في الحقوق والفرص بين الرجال والنساء، كما تحذر هيئة الأمم المتحدة للمرأة. (√)

Source Text	Translation Output
Therefore, it is of paramount importance to end the multiple forms of gender violence and secure equal access to quality education and health, economic resources and participation in political life for both women and girls and men and boys.	لذلك، من الأهمية بمكان إنهاء الأشكال المتعددة للعنف بين الجنسين وضمان المساواة في الحصول على التعليم الجيد والصحة والموارد الاقتصادية والمشاركة في الحياة السياسية لكل من النساء والفتيات والرجال والفتيان. (√)
It is also essential to achieve equal opportunities in access to employment and to positions of leadership and decision-making at all levels.	كما أنه من الضروري تحقيق تكافؤ الفرص في الوصول إلى الوظائف والمناصب القيادية وصنع القرار على جميع المستويات. (√)
The UN Secretary-General, Mr. António Guterres has stated that achieving gender equality and empowering women and girls is the unfinished business of our time, and the greatest human rights challenge in our world.	صرح الأمين العام للأمم المتحدة، السيد أنطونيو غوتيريس، بأن تحقيق المساواة بين الجنسين وتمكين النساء والفتيات هو عمل غير مكتمل في عصرنا، وأكبر تحد لحقوق الإنسان في عالمنا. (√)
The United Nations and women	الأمم المتحدة والمرأة. (√)
UN support for the rights of women began with the Organization's founding Charter.	بدأ دعم الأمم المتحدة لحقوق المرأة مع الميثاق التأسيسي للمنظمة. (√)
Among the purposes of the UN declared in Article 1 of its Charter is "To achieve international co-operation … in promoting and encouraging respect for human rights and for fundamental freedoms for all without distinction as to race, sex, language, or religion."	من بين مقاصد الأمم المتحدة المعلنة في المادة 1 من ميثاقها "تحقيق التعاون الدولي … في تعزيز وتشجيع احترام حقوق الإنسان والحريات الأساسية للجميع دون تمييز بسبب العرق أو الجنس أو اللغة أو الدين". (Fd)
Within the UN's first year, the Economic and Social Council established its Commission on the Status of Women, as the principal global policy-making body dedicated exclusively to gender equality and advancement of women.	في العام الأول للأمم المتحدة، أنشأ المجلس الاقتصادي والاجتماعي لجنته الخاصة بوضع المرأة، كهيئة عالمية رئيسية لصنع السياسات مكرسة حصريًا للمساواة بين الجنسين والنهوض بالمرأة. (√)
Among its earliest accomplishments was ensuring gender neutral language in the draft Universal Declaration of Human Rights.	وكان من بين أولى إنجازاته ضمان لغة محايدة جنسانياً في مشروع الإعلان العالمي لحقوق الإنسان. (√)

In table 8 below, the MT performance of Reverso Translation of Corpus 2 into Arabic is evaluated:

Table 8. Evaluating Reverso Translation Performance amid the existence of the input's parallel text in Arabic

Source Text	Translation Output
Gender Equality	(√) .المساواة بين الجنسين
The unfinished business of our time	(√) .الأعمال غير المنجزة في عصرنا
Women and girls represent half of the world's population and, therefore, also half of its potential.	وتمثل النساء والفتيات نصف سكان العالم، وبالتالي، تمثل أيضًا نصف إمكانياته. (√)
Gender equality, besides being a fundamental human right, is essential to achieve peaceful societies, with full human potential and sustainable development.	إن المساواة بين الجنسين، إلى جانب كونها حقًا أساسيًا من حقوق الإنسان، ضرورية لتحقيق مجتمعات مسالمة، ذات إمكانات بشرية كاملة وتنمية مستدامة. (√)
Moreover, it has been shown that empowering women spurs productivity and economic growth.	وعلاوة على ذلك، تبين أن تمكين المرأة يحفز الإنتاجية والنمو الاقتصادي. (√)
Unfortunately, there is still a long way to go to achieve full equality of rights and opportunities between men and women, warns UN Women.	وللأسف، لا يزال هناك طريق طويل يتعين قطعه لتحقيق المساواة الكاملة في الحقوق والفرص بين الرجل والمرأة، وهو ما تحذر منه هيئة الأمم المتحدة للمرأة. (x)
Therefore, it is of paramount importance to end the multiple forms of gender violence and secure equal access to quality education and health, economic resources and participation in political life for both women and girls and men and boys.	ولذلك، من الأهمية بمكان وضع حد للأشكال المتعددة للعنف الجنساني وضمان المساواة في الحصول على التعليم والصحة، والموارد الاقتصادية، والمشاركة في الحياة السياسية لكل من النساء والفتيات والرجال والفتيان. (√)
It is also essential to achieve equal opportunities in access to employment and to positions of leadership and decision-making at all levels.	ومن الضروري أيضًا تحقيق تكافؤ الفرص في الحصول على العمل وفي مناصب القيادة وصنع القرار على جميع المستويات. (√)
The UN Secretary-General, Mr. António Guterres has stated that achieving gender equality and empowering women and girls is the unfinished business of our time, and the greatest human rights challenge in our world.	وقد ذكر الأمين العام للأمم المتحدة، السيد أنطونيو غوتيريس، أن تحقيق المساواة بين الجنسين وتمكين النساء والفتيات هو العمل غير المنجز في عصرنا، وأكبر تحد لحقوق الإنسان في عالمنا. (√)

Source Text	Translation Output
The United Nations and women	الأمم المتحدة والمرأة. (√)
UN support for the rights of women began with the Organization's founding Charter.	وبدأ دعم الأمم المتحدة لحقوق المرأة بالميثاق التأسيسي للمنظمة. (√)
Among the purposes of the UN declared in Article 1 of its Charter is "To achieve international co-operation … in promoting and encouraging respect for human rights and for fundamental freedoms for all without distinction as to race, sex, language, or religion."	ومن بين مقاصد الأمم المتحدة المعلنة في المادة 1 من ميثاقها "تحقيق التعاون الدولي... وفي تعزيز وتشجيع احترام حقوق الإنسان والحريات الأساسية للجميع دون تمييز بسبب العرق أو الجنس أو اللغة أو الدين". (Fd)
Within the UN's first year, the Economic and Social Council established its Commission on the Status of Women, as the principal global policy-making body dedicated exclusively to gender equality and advancement of women.	وفي غضون السنة الأولى للأمم المتحدة، أنشأ المجلس الاقتصادي والاجتماعي لجنته المعنية بوضع المرأة، بوصفها الهيئة العالمية الرئيسية لصنع السياسات المكرسة حصرا للمساواة بين الجنسين والنهوض بالمرأة. (√)
Among its earliest accomplishments was ensuring gender neutral language in the draft Universal Declaration of Human Rights.	وكان من بين الإنجازات الأولى التي حققها ضمان صياغة محايدة جنسانيا في مشروع الإعلان العالمي لحقوق الإنسان. (√)

3.4 Analysis

3.4.1. Comparative Analysis of Translation Performance: GNMT vs. Reverso

Table 9 below uses translation errors per sentence as a scale to score translation errors of each performance of the aforementioned translation services:

Table 9. Scoring Google Translate and Reverso Translation Performances of corpus 2 in Arabic

Translation Service	Fluency Translation Error	Fidelity Translation Error	Correct Translation Score	Translation Error Percentage
Google Translate	1	0	13	3.57 %
Reverso Translation	2	1	12	10.71 %
Total	3	1	25	7.14 %

Figure 8 below shows the overall translation error percentages decrease from Corpus 1 (no parallel corpus used) to Corpus 2 (parallel corpus used) of PCBMT performances provided by Google Translate and Reverso Translation:

Figure 8. PCBMT error performance (percentage) comparison between corpus 1 and corpus 2 from English into Arabic
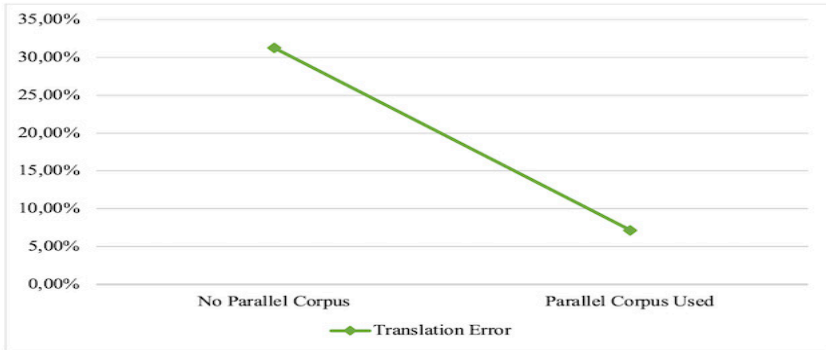
Despite Google Translate still suffering from problems on the fluency scale when translating from English into Arabic (scoring 1 error), it succeeded in providing a clean translation with only 3.57% of translation error. GNMT approaches, with the input's parallel text alread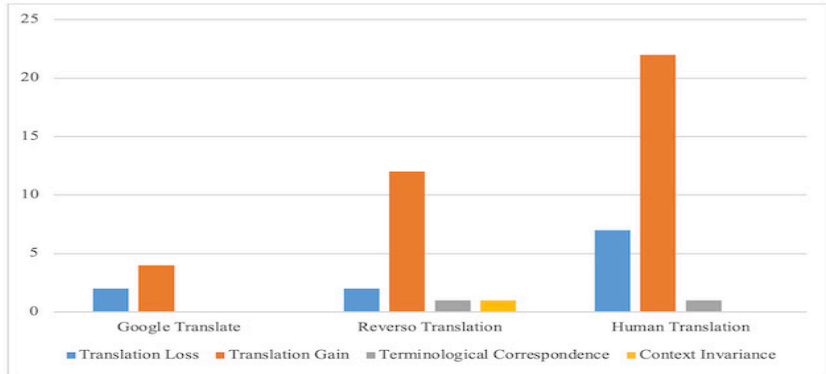y existing in Arabic, decreased the translation error percentage by 27.68%, providing a well-structured performance at both levels: fluency and fidelity. With zero (0) fidelity errors and thirteen (13) correct sentences, GNMT perfectly used the provided human translation to minimize errors at the most important levels for a translation to be well-performed. In a great success, GNMT provided six (6) out of eight (8) processed identical sentential expressions to the human translation, which points to the fact that NMT's probabilistic approach works well in such a situation.

Surprisingly, Reverso Translation provided a lower performance compared to GNMT in this translation directionality. Considering Reverso systems rely on SMT approaches, it was highly predictable that a better performance at both levels: fluency and fidelity, would be provided. However, Reverso's probabilistic approach failed at some key parts of the text processing. Scoring two (2) fluency errors and one (1) fidelity error, Reverso only decreased the translation error percentage to 10.71% (compared to 18.75% in Corpus 1), providing twelve (12) fully correct sentences. Despite this success, the SMT systems of Reverso Translation recalled only two (2) corresponding sentences

from the human translation, which again points to the SMT systems' (Reverso Translation case) probable inconsistency at some key levels of performance in such a situation.

Despite the disparity in PCBMT performances analyzed above, the overall translation error decreased from 25% in Corpus 1 to 7.14% in Corpus 2.

Tables 10 and 11 below show the correspondence errors of each PCBMT performance to Corpus 2 human translation in Arabic:

Table 10. Google Translate correspondence errors to the input's human translation in Arabic available online.

Source Text	Human Translation	Machine Translation
Translation and Gain Loss	/، بين، إلى، /، من، أمر، إطلاق، المجتمع، في، و، كما، هيئة،/، الإطار، الدولي، المعلن، /، الناس، على، ذلك، إطلاقا، لا، تفريق، الرجال، النساء، /، تحقيق	أيضا، بين، إلى، تنمية، مستدامة، من، /، /، /، /، /، /، /، /، /، التأسيسي، /، /، /، /، /، /، /، /، الخاصة، /.
Terminology	/.	تنمية مستدامة.
Context Invariance	No errors	No errors

Table 11. Reverso Translation correspondence errors to the input's human translation in Arabic available online

Reverso Translation			
Translation Correspondence Errors	**Source Text**	**Human Translation**	**Machine Translation**
Translation Loss and Gain	/, /, also, /, /, /, /, /, /, /, sustainable, development, has been, /, /, /, /, /, /, quality, /, /, /, founding, /, /, /, /, /, /, /, /, /, /, /.	بين، /، /، إلى، /، من، أمر، في، إطلاق، المجتمع، /، /، فقد، /، /، هيئة/، /، بين/، الجيد، الإطار، الدولي، المعلن، /، للناس، على، إطلاقا، لا، تفريق، الرجال، النساء، /، /.	بين، و، تمثل، أيضا، إلى، من، /، /، /، /، تنمية، مستدامة، /، يتعين، هو، ما، /، وضع، /، /، /، الإطار، /، التأسيسي، /، /، /، /، /، /، /، /، غضون، التي، حققها.
Terminology	Sustainable development, language	لغة /.	تنمية مستدامة، صياغة.
Context Invariance		No errors	1 error

Figure 9. Errors of both MT performances compared to HT

Interestingly, with the existence of an Arabic parallel text of the source text, Google Translate performed better than Reverso Translation. GNMT succeeded in recalling all terminological correspondences from the human translation, scoring zero errors in context invariance. Translation loss and gain errors were merely for the sake of accuracy in language structurality and text consistency. Reverso Translation, however, suffered from some key errors in both language structure and text coherence in particular. Its hybrid approach of rule-based + SMT strangely failed to recall some key structures and terminologies from the human translation.

The aforementioned points to the fact that NMT systems perform significantly better when a parallel text for the intended translation directionality is available (in this case, English-Arabic). The probabilistic recalling of NMT relatively had a persistent success in catching the corresponding word and term provided by the human translation, which was not the case for SMT represented by Reverso Translation. Therefore, it can be deduced that injecting web parallel corpora with quantities of human-translated texts from English into Arabic would strengthen the performance of PCBMT in this regard. The results provided by GNMT are truly encouraging and inspiring for further development of English-Arabic parallel corpora in terms of both quantity and quality.

3.4.2. Recalling Linguistic Appropriateness

While in Corpus 1 analysis, the focus was on evaluating the production of appropriate Arabic linguistics, in Corpus 2, it was more on evaluating how successful the PCBMT performance was in recalling appropriate Arabic linguistics from the humanly translated text.

3.4.2.1 Structural Consistency

Both PCBMT systems used in this study provided a cleaner performance for Corpus 2 compared to Corpus 1. However, instances of erroneous structural correspondences to the human translation were present. In Table 12 below, erroneous structural correspondences are listed together with the original structure and the human disposition in this regard.

Table 12. Erroneous structural correspondences to the Arabic human translation.

Source Structure	PCBMT System	PCBMT Output	Human Translation
Promoting and encouraging	GNMT/SMT	تعزيز وتشجيع	تعزيز احترام حقوق الإنسان والتشجيع على ذلك
As the principal global police making body	GNMT	كهيئة عالمية رئيسية لصنع السياسات	بصفتها الهيئة العالمية الرئيسية لصنع السياسات
Total of erroneous structural correspondences		2	

It is inappropriate in Arabic to juxtapose two grammatically undefined nouns in a sentence structure. Alternatively, Arabic prefers to have every undefined noun juxtaposed with its defining genitive. In the first example of erroneous structural correspondence shown in Table 11 above, the error appears despite the appropriateness of the human translation. The use of 'ك' before a word in Arabic is to express similarities between two things. The meaning here, though, was about the body itself as a principal global policy maker, not something resembling it.

3.4.2.2. Context

Both PCBMT systems used in this study provided a decent performance for Corpus 2 in Arabic. The already stored human translation of Corpus 2 in Arabic in their systems helped the performance achieve a certain degree of context invariance. GNMT provided quasi-perfect invariance in transmitting the intended contextual nuances from the source text, with a strong performance in recalling all necessary co-textual elements that express each sentence's context. SMT performance, represented by Reverso, was also stable in many parts, but a serious failure had an impact on context invariance. In the only erroneous sentence translation shown in Table 7 above,

a complete phrasal meaning was reversed. While the context was about the UN promoting women's rights, the sentence conveyed a meaning suggesting that the UN actually warns everyone about women. Such an error is clearly confusing for a reader new to this context and would directly attribute it to a translation error. As a result, GNMT has tangibly proven its outperformance in translating from English into Arabic when an input's parallel text is already available in the systems, both structurally and contextually.

Discussion and Conclusion

It is widely acknowledged that Arabic MT renditions need further research in MT studies to ensure output quality meets the requirements for minimal human revision. Indeed, PCBMT systems have propelled Arabic MT forward and moved away from the awkwardness prevalent in earlier approaches, providing Arabic renditions that meet basic requirements at many levels. However, quality issues persist in PCBMT Arabic performance, resulting in several errors in critical linguistic and contextual levels. Thus, this study addressed quality issues in two parts: first, the quality of the English-Arabic parallel corpus used by PCBMT systems, and second, the quality of these systems' outputs in the English-to-Arabic translation direction.

Based on the findings, this study was motivated by quality issues to quantitatively and qualitatively evaluate PCBMT system performance for English-to-Arabic translations. The methodology involved analyzing two PCBMT system performances with two different scenarios: one with an English source text and its human-translated parallel text in Arabic available online (Corpus 2), and another without such a parallel text available online (Corpus 1). Both source texts addressed the same theme, ensuring similar contextual meaning. The evaluation included a hybrid of comparative and contrastive analyses between Corpus 1 and Corpus 2 for two different PCBMT systems: NMT represented by Google Translate and SMT represented by Reverso Translation. Errors in PCBMT output sentences from both corpora were scored on two main scales: fluency and fidelity, to measure overall translation quality. A quantitative analysis followed to gauge the quality of each performance independently.

The dual-situated analysis highlighted divergent PCBMT system performances for Corpus 1 and Corpus 2 in Arabic, demonstrating sensitivity in recognizing PCBMT performance variations in both corpus situations. Thus, the evaluation moved beyond ordinary PCBMT performance analysis to include an assessment of how well each PCBMT system recalled the

existing human translation of the same text. The results gained particular clarity through comparative and contrastive analyses at translational and linguistic levels alike.

As previously mentioned, the results of Corpus 1 and Corpus 2 PCBMT performances diverged significantly in terms of lexicon, sentence structure, terminology, and contextual invariance across the four performances evaluated in this study. The finding that GNMT systems provided better performance for Corpus 2 than Reverso SMT systems is an intriguing discovery that warrants further investigation. The results also underscore that while PCBMT performances in other translation directions are increasingly focused on stylistics, issues persist when evaluating English-to-Arabic performance in MT errors. The analysis of Corpus 1 clearly illustrates the challenges of Arabic as a target language in MT, despite English being a common source text for other languages and MT systems being customized accordingly. Arabic renditions still suffer from core errors in language structure and contextual invariance, even with the use of PCBMT systems and the aim of recalling corresponding structures and contextual nuances from refined human translations.

In conclusion, this study is limited in scope, as it is impractical to evaluate all PCBMT renditions of English into Arabic. Nevertheless, this exploratory study provides a concrete analysis of PCBMT performance variations between Corpus 1 and Corpus 2 despite their shared thematic context. Despite its limitations, this study suggests further evaluations and analyses of PCBMT systems in diverse contexts and scenarios, particularly in the context of English-Arabic translation directionality. Additionally, evaluating the reverse directionality (Arabic-to-English) would also be of particular interest.

Bibliography

Alkahtani, S & William, J. T. (2015). Building and verifying parallel corpora between Arabic and English. *(Doctoral thesis, Bangor University).* Bangor University theses. https://research.bangor.ac.uk/portal/files/20574012/null

Alsoheybe, N. T., Dahan, N. A., & Ba-Alwi, F. M. (2017). Machine-translation history and evolution: Survey for English-Arabic translations. *Current Journal of Applied Sciences and Technology, 23* (4), 1-19. 10.9734/CJAST/2017/36124

Al-Sulaiti, L., & Atwell, E. (2006). The design of corpus of contemporary Arabic. *International Journal of Corpus Linguistics, 11* (2), 135-171. 10.13140/2.1.2228.8320

Ashengo, Y, A., & Aga, R. T., & Abebe, S. L. (2021). Context Based Machine Translation with Recurrent Neural Network for English-Amharic Translation. *Machine Translation. 35,* 19-36. https://doi.org/10.1007/s10590-021-09262-4

Baker, Mona. (1995). Corpora in translation studies: An overview and some suggestions

for future research. *Target, 7* (2), 234–243. DOI 10.1075/target.7.2.03bak

Baldwin, T & Kordon, V. (2011). The interaction between linguistics and computational linguistics. *LiLT, 6* (1), 1–6. https://journals.linguisticsociety.org/elanguage/lilt/article/download/2577/2577-5346-1-PB.pdf

Catford, J. C. (1965). A linguistic theory of translation: An essay in applied linguistics. Oxford: Oxford University Press.

De Beaugrande, R, & Wolfgang, D. (1981). *Introduction to Text Linguistics*. London: Longman.

Dillon, M. A., Hojeij, Z., Perkins, A., & Malkawi, R. (2018). Examining the test quality of English/Arabic dual language children's picture books. *International Journal of Bilingual Education and Bilingualism, 23* (8), 888–901. https://doi.org/10.1080/13670050.2017.1415867

Dogru, G., Martín-Mor, A., & Aguilar-Amat, A. (2018). Parallel Corpora Preparation for Machine Translation of Low-Resource Languages: Turkish to English Cardiology Corpora. *At the conference of Proceedings of the LREC 2018 Workshop 'MultilingualBIO: Multilingual Biomedical Text*. https://www.researchgate.net/profile/Gokhan-Dogru/publication/337331602_Parallel_Corpora_Preparation_for_Machine_Translation_of_Low-Resource_Languages_Turkish_to_English_Cardiology_Corpora/links/5dd2b3bf92851c382f499a3b/Parallel-Corpora-Preparation-for-Machine-Translation-of-Low-Resource-Languages-Turkish-to-English-Cardiology-Corpora.pdf ? origin =publication_detail

Dorr B & Olive J & McCary J & Christianson, C. (2011). Machine Translation Evaluation and Optimization. In: Olive J & Christianson C & McCary J. (eds). *Handbook of Natural Language Processing and Machine Translation* (745–843). Springer, New York, NY. https://doi.org/10.1007/978-1-4419-7713-7_5

Dorr, J. B. (1994). Machine translation: A view from the lexicon. *Computational Linguistics, 20* (4), 670-676. https://aclanthology.org/J94-4009.pdf

Doval, Irene & Sanchez, Teresa. (2019). *Parallel Corpora for Contrastive and Translation Studies*. Amsterdam: John Benjamins Publishing.

Fiederer. R., & O'Brien, S. (2009). Quality and machine translation: A realistic objective?. The Journal of Specialized Translation, (11), 52-74.

Forcada, Mikel. L. (2017). Making sense of neural machine translation. *Translation Spaces, 6* (2), 291–309. doi 10.1075/ts.6.2.06for

Frankenberg-Garcia, A. (2008). Suggesting rather special facts: A corpus-based study of distinctive lexical distributions in translated texts. *Corpora, 2* (3), 195–211. https://doi.org/10.3366/E1749503208000154

Frankenberg-Garcia, A. (2021). Can a corpus-driven lexical analysis of human and machine translation unveil discourse features that set them apart. *Target*, 1-33.

Halliday, M. A. K. (1978). Language as a Social Semiotic: The Social Interpretation of Language and Meaning. London: Edward Arnold.

House, J. (2006). Text and Context in Translation. *Journal of Pragmatics, 38* (3), 338–358. https://doi.org/10.1016/j.pragma.2005.06.021

Hovy, E. H., Margaret, K., & Popescu-Belis, A. (2002). Computer-aided specification of quality models for MT evaluation. *In LREC 2002: Third International Conference on Language Resources and Evaluation*, Las Palmas de Gran Canaria, Spain, 1239–1246.

Hovy, Eduard. (2002). Principles of context-based machine translation evaluation. *Machine translation, 17*, 43 – 75. http://www.cs.cmu.edu/~./hovy/papers/03MT-FEMTI-MT-eval.pdf

Khadim, K. A. (2014). A stylistic analysis of news English-Arabic translation. *Social Sciences and Humanities, 22* (3), 797–823.

Nizar, H., Nasser, Z., Dima, T., Hieu, H., & Maverick, A. (2017). A parallel corpus for evaluating machine translation between Arabic and European languages. *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, Valencia, Spain, 235–241. https://aclanthology.org/E17-2038.pdf

Peter, S & Jean, O. (2013). Theory and applications for natural language processing: Essential speech al language technology for Dutch. Heidelberg: Springer. doi 10.1007/978-3-642-30910-6

Popovic, Maja. (2017). Comparing language related issues for NMT and PBMT between German and English. *The Prague Bulletin of Mathematical Linguistics, 108,* 209–220. 10.1515/pralin-2017-0021

Resnik, Philip, & Smith, Noah A. (2003). The Web as a Parallel Corpus. *Computational Linguistics. 29* (3), 349-380. 10.1162/089120103322711578

Sinclair, Jean. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Sparck, J. K, & Galliers, J. R. (1996). Evaluating natural language processing systems: An analysis and review. Berlin: Springer-Verlag.

Vaida, Pragati. (2014). Decoding in statistical machine translation using Moses and Cygwin on windows. *International Journal of Engineering Research & Technology, 3* (2), 736 – 74. https://aclanthology.org/W08-0510.pdf

Van Dijk, T A. (1977). Text and context: Explorations in the semantics and pragmatics of discourse. Harlow: Longman.

Vinay, J. P & Darbelnet, J. (1958) (2004). *A methodology for translation*. In The Translation Studies Reader (2nd ed), edited by Lawrence Venuti. London: Routledge.

Wei, H & Yushan, Z. (2014). A corpus based machine translation method of term extraction in LSP texts. *Theory and Practice in Language Studies, 4* (1), 46-51. doi:10.4304/tpls.4.1.46-51

Zanettin, Federico. (2011). Translation and corpus design. *SYNAPS – A Journal of Professional Communication, 26*, 14–23. https://openaccess.nhh.no/nhh-xmlui/bitstream/handle/11250/2394011/Zanettin_26.pdf?   sequence   =1&isAllowed =y

Zhou, D & Wang, Y. (2015). Corpus-based machine translation: Its current development and perspectives. *International Forum of Teaching and Studies, 11* (1-2), 90–95.

Appendix 1: Links of corpora analysed and examples used

Corpus   1:   https://au.int/en/documents/20200207/annual-report-activi-

ties-union-and-its-organs-2020
Corpus 2: English text: https://www.un.org/en/global-issues/gender-equality
Arabic human translation: https://www.un.org/ar/global-issues/gender-equality
Example 1 (figure 1): https://opus.nlpl.eu/TED2013/v1.1/ar-en_sample.html
Example 2 (figure 2): https://opus.nlpl.eu/infopankki/v1/ar-en_sample.html

Abstract

This article aims at providing a practical quantitative analysis to evaluate parallel corpora-based machine translation (PCBMT) performances using an English into Arabic directionality. This article proposes to compare PCBMT performances of two different source texts in two different situations. The first situational PCBMT performance evaluation embarks from the non-existence of an online Arabic parallel text to the source text. The second situational PCBMT performance evaluation, however, uses another source text which has a human-translated parallel text available online for public use. Subsequently, a quantitative analysis is held to gauge quality in both performances in light of their different situations. Upon this, the present study explores the extent to which PCBMT systems' performance varies between situation 1 and situation 2, in terms of observing Arabic structural and contextual requirements to provide the ultimate output. The study shows that PCBMT performance from English into Arabic is quite divergent between the two situations and reiterates the need to further investigate Arabic renditions of PCBMT performances in both directionalities, from and into Arabic.

Keywords

Machine translation evaluation, parallel corpora, PCBMT performance, structurality, context

Résumé

Cet article vise à fournir une analyse quantitative pratique pour évaluer les performances de la traduction automatique basée sur des corpus parallèles (TABCP) de l'anglais vers l'arabe. Cet article propose de comparer les performances de TABCP de deux textes sources différents dans deux situations différentes. La première évaluation situationnelle de la performance du TABCP entame de l'inexistence d'un texte parallèle arabe en ligne au texte source. La deuxième évaluation situationnelle de la performance du TABCP utilise cependant un autre texte source dont la traduction humaine est disponible en ligne pour un usage public. Par la suite, une analyse quantitative est effectuée pour évaluer la qualité des deux performances à la lumière de leurs différentes situations. La présente étude examine la mesure dans laquelle la performance des systèmes TABCP varie entre la situation 1 et la situation 2, en ce qui concerne l'observation des normes structurelles et contextuelles de la langue

arabe en fournissant le texte final. L'étude montre que la performance du TABCP de l'anglais vers l'arabe est très divergente entre les deux situations et réaffirme la nécessité d'approfondir les performances en arabe du TABCP dans les deux directions, de et vers l'arabe.

Mots-clés

Évaluation de la traduction automatique, corpus parallèles, performance du TABCP, structuralité, contexte

مستخلص

يهدف هذا المقال إلى تقديم تحليل كمي وعملي لتقييم أداء برامج الترجمة الآلية القائمة على المدونات المتوازية باستخدام اتجاه تُرجمي إنجليزي عربي. إذ تقترح هذه الدراسة مقارنة أداء هذه البرامج عند ترجمة نصين مختلفين في وضعيتين مختلفتين. وينطلق تقييم الأداء في الوضعية الأولى من حقيقة عدم وجود نص موازٍ باللغة العربية متاح على الإنترنت للنص الأصلي باللغة الإنجليزية. بينما تقييم الأداء في الوضعية الثانية ينطلق من نص إنجليزي آخر نصه الموازي باللغة العربية متاح على الإنترنت للاستخدام العام. بعد استخراج الأداء، تقدم هذه الدراسة تحليلاً كمياً لقياس مدى جودة الأداء في البرنامجين على ضوء اختلاف وضعيتهما. على إثر ذلك، تغوص هذه الدراسة في مدى اختلاف أداء برنامجي الترجمة الآلية المعتمدة بين الوضعيتين الأولى والثانية، فيما يتعلق باحترام المعايير البنيوية والسياقية للغة العربية عند تقديم المنتج النهائي. وقد أظهرت هذه الدراسة أن أداء برنامجي الترجمة الآلية القائمة على المدونات المتوازية من اللغة الإنجليزية إلى اللغة العربية قد اختلف اختلافاً كبيراً بين الوضعيتين، وتؤكد الدراسة على الحاجة المُلحة إلى التقصي أكثر في النتاج العربي لأداء هذه البرامج وكذلك في الاتجاهات التُرجمية الأخرى من وإلى اللغة العربية

كلمات مفتاحيّة

تقييم الترجمة الآلية، المدونات المتوازية، أداء برامج الترجمة الآلية القائمة على المدونات المتوازية، البنيوية، السياق