Scientific Research Publishing

Open Journal of Modern Linguistics, 2020, 10, 524-548
https://www.scirp.org/journal/ojml
ISSN Online: 2164-2834
ISSN Print: 2164-2818

Quality and Machine Translation: An Evaluation of Online Machine Translation of English into Arabic Texts

Mohammed Abdulmalik Ali

Department of English, Prince Sattam bin Abdulaziz University, Al-Kharj, KSA
Email: mhajali2000@gmail.com

How to cite this paper: Ali, M. A. (2020). Quality and Machine Translation: An Evaluation of Online Machine Translation of English into Arabic Texts. *Open Journal of Modern Linguistics, 10,* 524-548.
https://doi.org/10.4236/ojml.2020.105030

Received: August 2, 2020
Accepted: October 12, 2020
Published: October 15, 2020

Abstract

This study compares the translation outputs of an English into Arabic text using the three machine translators of Google Translate, Microsoft Bing, and Ginger. To carry this evaluation of the machine translation (MT) outputs, an English text and its Arabic counterpart were selected from the UN records. The English source text was segmented into 84 semantic chunks. Depending on the Arabic counterpart model text, each chunk was rated as "correct or incorrect" at the two levels of the translation attributes: fidelity and intelligibility. To perform the quantitative description of the evaluation process, the numbers of fidelity and intelligibility errors and their percentages were calculated. Results of this evaluation process revealed that none of the three translated versions of the source text was perfectly translated. Although the translation of Microsoft Bing was rated the best, Google's translation was found the least accurate due to the high percentage of fidelity and intelligibility errors detected in its translation output. However, the quality of Ginger's translation was found slightly less accurate than that of Microsoft Bing, but remarkably better than Google's translation. The findings of this study imply that these MT applications can be implemented to perform English into Arabic translation to get the broad gist of a source text, but a deep and thorough post-editing process looks essential for a full and accurate understanding of an English into Arabic MT output. The study recommends that more studies are encouraged to continue to assess the quality of MT that will further highlight its weaknesses and the strategies that should be adopted to overcome them.

Keywords

Machine Translation, Quality of Machine Translation, Online Translator, Translation Errors

1. Introduction

Machine translation (MT) which is a subgroup of computational linguistics is defined as "the process that utilizes computer software to translate text from one natural language to another" (Alawneh & Sembok, 2011: p. 343). The use of MT on PCs and smartphones "has become increasingly widespread in various settings because of its convenience, multilingualism, immediacy, efficiency, and free cost" (Lee, 2019: p. 158), regardless of some of its deficiencies. In general, the process of human translation usually starts by deciphering or decoding the meaning of a source text (ST), then moves to re-encode this meaning in the target text (TT) of another language, post-editing or evaluation of the output comes as a final stage to make sure that the message presented in the ST has been conveyed effectively (Cieślak, 2011; Wilss, 1982). To decode the meaning of the ST, a translator should interpret and analyze all of its components, a process that needs in-depth knowledge of the grammar, syntax, semantics, morphology, etc., of the SL. The same in-depth knowledge, if not deeper knowledge of the target language (TL) is needed for the second step of re-encoding the meaning, and the third step of evaluating the output TT (Doherty, 2016).

In the same sequence, Zong (2018) explains that the process of machine translation relies on the analysis of words, grammar, meaning, and style. This MT process starts by dividing the sentence into words, followed by identifying the meaning of each word through the online dictionary, then by analyzing the sentence or clause according to the followed grammar rules in order to convert it into a conceptual construct, and finally, a target language model is used to generate the sentence or text in the TL. The language model according to Zong (2018) "is the intermediate language between the source language (SL) and the TL, through which various languages can be translated into another desired language. If coupled with bidirectional translation software, the automatic translation system can translate multiple languages." (p. 4).

Translation professionals stated that the main parameter for evaluating MT systems is the quality of their translated outputs (Zong, 2018). Zong (2018) clarifies that with the feature of deep learning abilities that are developed through artificial intelligence, the current translation systems are gradually improving the quality of MT outputs. However, no one dares to claim that MT outputs are perfect, despite these successive improvements in MT at this artificial intelligence era as reflected in the neural approach of MT. Therefore, it is argued that there are still some translation problems that are far from being fully resolved, such as the wrong choice of words, the spelling of words, and the translation of words and sentences out of their contextual sequence. Consequently, human proofreaders are always needed to revise the MT outputs to correct errors and improve their quality (Ali, 2016, 2018). To reduce these human efforts and interferences, researchers are even working on developing an automated evaluation of computer translation systems. Among the efforts exerted to improve these evaluation processes, White (2001) has attempted to formulate methods to

automatically evaluate machine translation outputs through revising some of the translation attributes, and then tried, explicitly or implicitly, to extrapolate the measurement to cover a broader class of attributes. Some studies have focused on measuring fidelity and intelligibility as main attributes of translation in general, and of MT in particular as is the case in the present study (e.g., Taleghani, & Pazouki, 2018; Fiederer & O'Brien, 2009). Yet, regardless of the different evaluation factors, researchers argue that the "quality" of translation outputs has always been treated as the critical key to indicate the effectiveness of machine translation.

To overcome translation difficulties, White (2001) as well as other earlier researchers like Hirschman, Reeder, Burger, and Miller (2000), and Lee (2019) worked on operating some standards for carrying out "automated MT evaluation". They argue that methods of automated MT evaluation depend on two main attributes of translation: fidelity and intelligibility. While fidelity (also called accuracy) is meant to measure the "conveyance of the information in the source expression into the target expression", intelligibility (also called fluency) measures "how understandable the target expression is to a target-native speaker" (White, 2001: p. 1). White (2001) also explains that these attributes are measured by comparing MT outputs to models of co-occurrences of that TL. In addition to these two attributes, Hutchins and Somers (1992: p. 163) tested "style" as a third attribute to evaluate MT outputs. Style is the translation attribute wherein the extent to which a TT uses the appropriate language to convey the message effectively. Other researchers argue that although "style" could be a factor in the assessment of translated materials, it does not hinder the conveyance of the message in the TT. This has been emphasized by Bowker, (2014), who revealed that many—though not all—of the participants of his study were satisfied with texts that are semantically accurate, but which need not be stylistically elegant. Additionally, other researchers like Fiederer and O'Brien (2009), and Taleghani and Pazouki (2018) have investigated the quality of MT outputs depending on the fidelity and intelligibility attributes without considering the style as a quality factor. To evaluate MT outputs, Doherty (2016) uses the terms accuracy (for fidelity) and fluency (for intelligibility). To him, accuracy "denotes the extent to which the meaning of the source text is rendered in its translation, and fluency denotes the naturalness of the translated text in terms of the norms of that language" (958). Yet, Doherty claims that assessing translation quality and telling what is good or not had always been debated compared to human translation, and MT has deepened that debate, but this debate is leading to the progress in MT and automatic evaluation to cope with human translation on the one hand, and the evaluation of translation outputs, on the other hand.

In this sequence of studying the effectiveness of MT depending on the fidelity and intelligibility attributes of its outputs, the present study arose to assess the quality of machine translation of an English text (ST) into Arabic (TL) using three free online translation programs: Google Translate, Microsoft Bing, and

Giger. This study also aimed to determine which MT application among the three chosen ones produces an Arabic TT of a higher quality which would be more appropriate for Arabic native speakers to understand straightforwardly.

2. Significance of Study

The rapid increase of globalization and the wish to penetrate new markets, the need for Arabic native speaking people to get in touch with other scientific and cultural matters written in English, the on-going need to reduce the cost of translation, and the requirement to publish translated materials at the same time as source language material, are all vital reasons for conducting this kind of study as an effort to evaluate MT outputs regarding some chosen translation systems. Additionally, some researchers (e.g., Ali, 2018; and Doherty, 2016) argue that there is a shortage of qualified translation professionals, and so a growing demand for effective MT to compensate for this shortage looks crucial. Moreover, although the literature on MT is vast, there is a marked shortage of machine translation studies in the field of MT related to translation from English to Arabic as a TL. Native speakers of Arabic usually attend to translate English texts into Arabic or vice versa using MT systems despite the awareness of their deficiencies; and this could be because they may need to save time, effort, and/or cost. Yet, some MT applications can produce more accurate translations than others (Albat, 2012). It is also obvious that the quality MT outputs vary from one application to another, and from one language to another (Fiederer & O'Brien, 2009), and so the present study aimed to investigate the quality of translation from English into Arabic using three different machine translation applications. This comparative type of study would help users be aware of the quality of the English into Arabic translated materials that are performed by different translation applications. Hence, results of this investigation are expected developers of MT systems become aware of the customers' needs and the shortcomings they should try to overcome. Inspired by these circumstances, this study deemed to evaluate the translation of an English into Arabic text of three MT tools. Therefore, an English text known to be written in a formal standard language by native speakers of English would be chosen for translation into the Arabic language. created in MS Word 2007, provides authors with most of the formatting specifications needed for preparing electronic versions of their papers. All standard paper components have been specified for three reasons: 1) ease of use when formatting individual papers, 2) automatic compliance to electronic requirements that facilitate the concurrent or later production of electronic products, and 3) conformity of style throughout a journal paper. Margins, column widths, line spacing, and type styles are built-in; examples of the type styles are provided throughout this document and are identified in italic type, within parentheses, following the example. Some components, such as multi-leveled equations, graphics, and tables are not prescribed, although the various table text styles are provided.

3. Literature

3.1. Machine Translation Approaches and Development

Machine translation has gone through different phases that led to the present situation where users can deal with its translated outputs with some kind of confidence for some languages. Three approaches to MT have arisen since the 1980s due to the increasing spread of computer applications. At the early stages of MT, the rule-based approach in which linguistic rules were written by professionals for the source and target languages was the most dominant (Doherty, 2016). These rules initially prescribed the morphological and syntactic rules and provided the semantic analysis of each language pair to form what is called Translation Memory (TM).

Later, and depending on the accumulative TMs, a shift from the prescriptive (rule-based) approach to the descriptive approach (called statistical approach) took place. To describe how the statistical MT approach works, Doherty (2016) reports that it uses "complex statistical algorithms to analyze large amounts of data to generate a monolingual language model for each of the two given languages, and a translation model for the translation of words and phrases from one of these languages into the other. A decoder then uses these models to extrapolate the probability of a given word or phrase being translated from one language into the other, where the most probable word or phrase co-occurrences are chosen as the best translation." (p. 953)

Enhanced by the development of artificial intelligence, the newest approach of the neural MT has taken place recently. To describe it, Zong (2018) states that the neural MT achieved high quality because it is characterized by using the neural networks, where it continuously receives, in the backstage, different training data to perform data mining and training through deep learning capabilities. Zong (2018) also adds that this approach is based on the theory and techniques of natural language understanding, natural language processing, machine translation, translation memory, and statistics-based machine translation as well as deep learning.

3.2. Previous Research

Closely related to the present study, Taleghani, and Pazouki (2018) conducted a study to evaluate and compare the translation of English idioms and phrasal verbs into Persian using the four free online translation systems: (www.bing.com, www.translate.google.com, www.freetranslation.com, and www.targoman.com). The focus was on the translation of idioms and phrasal verbs occurring in ten English texts. The translations of the targeted idioms were subjectively compared to their equivalents in a Persian dictionary. Results of the comparisons showed that the translation of "www.targoman.com" was of better quality than the other three systems, and so the study recommended this system for translating idioms and phrasal verbs from English to Persian.

Similarly, in an evaluation study of MT, Al-Khresheh and Almaaytah (2018)

used Google Translate to translate separate English proverbs into Arabic. The study revealed that "Google translate" faced some linguistic obstacles in conveying the same meanings of English proverbs into Arabic because of word ambiguity (polysemy and homonym), on the one hand. On the other hand, the different systems of syntax and sentence structures in English and Arabic made the process of translation sometimes challenging.

In a more recent study, Daniele (2019) quantitatively assessed the performance of a free online translator in translating medical texts from English into Italian. Translation effectiveness was evaluated and established by analyzing the number and percent of errors occurring in the translation of original research abstracts in the medical field. Furthermore, this study analyzed the total number and the percent of translation errors and their correlation with lexical density. The mean percentage of total translation errors was 15%. A direct correlation was also found between the total translation errors and the lexical density. The findings indicated a fairly good performance of Google Translate in translating words in highly academic writings such as medical abstracts. The study concluded that an effective translation is not only a matter of finding a correspondence between words in the source language and the target language; many other aspects are just as important.

Fiederer and O'Brien (2009) studied the quality of MT output of English into German sentences using IBM WebSphere as the MT system. Eleven qualified specialists rated 30 source sentences distributed into three translated versions and three post-edited versions for the attributes of clarity, accuracy, and style. Findings of the study revealed that the MT sentences that were post-edited were rated to be of higher clarity and accuracy, while the human translations were judged to be of better style. However, when the evaluators were asked to choose the type of translated sentences they preferred, the majority chose the human translated sentences. However, in Bowker's study (2014), which investigated the potential of machine translation for aiding Spanish-speaking newcomers to Canada to make better use of the Ottawa Public Library's (OPL) Website, revealed that many of the participants were satisfied with human and the post edited MT outputs that were semantically accurate, but which did not need be stylistically elegant.

The belief of some researchers in MT inaccurate outputs, motivated translation instructors and researchers to make use of those outputs in classroom instruction to improve their foreign language learners' translation abilities and language skills. In this regard, literature has shown researchers' interest to investigate the use of post-editing of MT raw materials for enhancing undergraduate F/SL students' language competencies and their translation capabilities. Niño (2009), for example, studied the effect of the "post-editing" technique within the evaluation process of MT outputs on enhancing learners' foreign language abilities. The participants' task in the study was to make all corrections needed to produce readable texts out of the raw MT outputs. Findings of this study as well as those similar (e.g., La Torre, 1999; Niño, 2004; Belam, 2003; and Kliffer, 2005)

showed different uses of MT outputs in foreign language contexts, and the usefulness of implementing such materials for evaluation and post-editing purposes in enhancing learners' language competencies. Not far from this type of study, Belam (2003) implemented the MT evaluation strategy in an introductory MT course for undergraduate students. Students were asked to design a project to evaluate MT tools or to conduct a comparative evaluation of MT texts of different types. Belam reported different benefits of the experiment on students' language and translation abilities. Similarly, Kliffer (2005) used post-editing in an undergraduate course of French to English translation. Positive results were reported, among them was that post-editing of MT texts made the translation process less stressful for students than performing the entire translation on their own.

4. Research Questions

This study aimed to evaluate and compare the effectiveness of three online translators in translating an English language text taken from the UN records into Arabic, and to find out the translator which would produce the best translation. More precisely, the study aimed to answer the following questions:

1) How effective are the online MT applications in translating English into Arabic texts?

2) Which MT application results in a more accurate translation of English into Arabic texts?

5. Methods

This comparative study aimed to evaluate the quality of MT outputs by comparing the numbers and percentages of errors occurring in English into Arabic translation outputs using three MT applications. The quantitative analysis of the numbers of errors related to the translation attributes of fidelity and intelligibility would lead to the identification of the MT application among the three used ones that would result in a higher quality of the translation of an English into Arabic text. Each one of the two languages targeted belongs to different language families that have diverse linguistic systems and cultures. While English belongs to the Germanic family of languages, Arabic belongs to the Semitic family (Al-Khresheh, 2016). This makes the translation process a bit complicated and sometimes results in numerous translation errors that distort the concept to be conveyed. To achieve the main objectives of this study and to discover the quality of MT, an English text was selected, then translated by three MT systems. A descriptive qualitative and quantitative analysis of the three outputs was carried out depending on the researchers' native language intuition (Arabic) and his linguistic background of English and Arabic.

5.1. Materials

5.1.1. Machine Translation Applications

To choose the appropriate online translators of English texts (ST) into Arabic

(TT), it appeared reasonable to select the most commonly and frequently used applications by Arabic native speakers. A list of packages was shown on the Google searching engine as a result of entering the keywords (free + online + translate + English + Arabic). The investigation of the many applications listed included Google Translate as the first option since it is the most commonly used system for a quick translation, and the second application appearing in the list was Microsoft Bing. These applications were also said to be the most commonly used applications for translation purposes (Doherty, 2016). Among the applications appearing in the list, some were excluded because their Arabic TT outputs looked remarkably unconvincing such as "Yandex": https://translate.yandex.com/translator/English-Arabic. Others were excluded as they were available for commercial purposes, such as http://www.easyhindityping.com/english-to-arabic-translation. A third category offered the translation of texts of no more than 200 words, such as https://www.systransoft.com/lp/arabic-translation/, and so, they were excluded. Going through the inspection of the list, helped the researcher to choose a third application called "Ginger" which was also expected to be commonly used in the Arabic context in the Middle East at least because Arab users could be encouraged to use it not only to translate English texts but also to paraphrase, check their English written work and for other services. The inspection of long list MT applications ended by choosing the following three translators because at least their translations seemed to be more convincing than others, and they had the feature of translating texts above 200 words. (Google Translate: https://translate.google.com.sa/?hl=en&tab=rT)

Google supports translation for more than 50 languages. When users access this application, they see a single page. The application offers the options to write, paste, or upload a text or a document that does not exceed 5000 characters. This software can translate from more than 70 SLs into the same number of TLs. The languages into which the package translates are listed clearly on the screen. The translated output can be shared on Twitter or via email and it can be copied. The user can even listen to that TT read by the system. The layout of this package is clear and highly user-friendly, with no adverts for other products and no unnecessary written texts which may confuse the screen. However, the screen does not have an icon to help the user print the TT directly. (Microsoft Bing: https://www.bing.com/translator)

The Microsoft Bing translator gives its users the option to key a text or webpage URL, with no indication of ST word limit. Similar to the Google Translate application, it is clear that Microsoft Bing can translate from 70 SLs into the same number of TLs. It is not possible to print the TT directly from the screen nor to speak it although a speaker icon is available on the screen, which may be inconvenient. However, the layout of the package is clear, and it is user-friendly. (Ginger: https://www.gingersoftware.com/translation)

Ginger is an application that requires the user to download, and it has a free and a paid version. It provides services for managing content and language

translation. When accessing Ginger, users are given the option to enter an ST, but with no word limit. The screen shows that it can translate from more than 70 SLs into the same number of TLs. It is not possible to print the TT directly from the screen nor to speak it although a speaker icon can be seen on the screen, a case which may look inconvenient. However, the layout of the package is clear, but less user-friendly compared to the other two packages tools. Furthermore, this application has also several options like "speaker, paraphrase, write, and language checker", and these options could be some of the reasons that motivated students to use.

5.1.2. Choice of Source Text

White (2003) argues that since translating a given ST can be done in different correct ways, then different ways could be adopted to evaluate the translated texts. This may lead to some kind of subjective evaluation. Additionally, professional translators and evaluators may have different reactions to the subject matter of the analysis of a translated text the second time they see it than they do the first time (White, 2003: p. 218). This fact inspired the writer of this article to choose an English ST that has a formal TT (Arabic) counterpart to avoid any subjectivity in the evaluation process. Another criterion to choose the English ST to translate into Arabic was that it should be written by proficient native speakers of English to have an error-free text. To have a text as such, it was decided to choose it from the UN records as a traditional documentary written work which, like all UN documents, has its counterparts in six standard languages including Arabic. This way, the UN Arabic copy of the text would be treated as the model or reference to compare the MT translation outputs with.

To investigate the effectiveness of machine translation, and to make sure that the applications chosen were given an English text of the same language complexity and subject matter (theme), a text talking about the financial, budgetary, and administrative matters from the "Annual Session 2020" report of the "Executive Board of the United Nations Development Programme", was chosen, (**Appendices 1**, and **2**):

- English ST:
https://digitallibrary.un.org/record/3861578/files/DP_2020_9-EN.pdf, (**Appendix 2**).

- Arabic counterpart TT:
https://digitallibrary.un.org/record/3861578/files/DP_2020_9-AR.pdf, (**Appendix 3**).

It is also important to highlight that a text of about 344 words long was appropriate for increasing the reliability of the MT evaluation measure (Turian, Shen, & Melamed, 2003).

5.1.3. Evaluation Method of Target Texts

There are different methods to evaluate MT systems and their outputs. The oldest is the use of human evaluators to assess the quality of translated materials (Anderson, 1995). Although human evaluation is time-consuming, it continues

to be the most reliable method to compare different systems using the rule-based, or the statistical translation approaches (Han, Wong, & Chao, 2012), or the recently developed neural approach. To investigate the effectiveness of MT in the English Arabic context, a comparative evaluation of the quality of the translation output which is based on the adequacy of the translated text "or its fidelity", and the fluency of the translated text "or its intelligibility" (White, 1995; Doherty, 2016). The study opted to conduct the human evaluation type of the quality of the MT outputs rather than the automated method. This process required a descriptive analysis of the numbers and means of correct and incorrect ratings of the translated segments or chunks. This rating process would be based on comparing the translated chunks to the reference (model) counterpart text (formal Arabic UN TT), because "the closer a machine translation is to a professional human translation, the better it is" (Papineni, Roukos, Ward, & Zhu, 2002: p. 311). Translation quality for each of the output texts would be calculated through the total numbers of fidelity and intelligibility errors and the percentage of total translation errors related to each MT application, and this would eventually lead to identifying the application which would perform a higher quality of the English into Arabic translation.

The present study opted to use the "Black box" approach to analyze the MT translated materials because it aimed to compare the ST as an input and the TT as the output of the translation process without considering the mechanics of the translation engine. Applying this approach, or what White (2003) called the "declarative evaluation" requires evaluating the extent to which the TT is "faithful" to the information conveyed in the ST, and the extent to which fluency is maintained in the TT. In fact, determining the degree of fidelity and intelligibility is an essential step in the process of translation, whether carried out by machines or humans.

Quoting White's (2001) method of MT evaluation, and to carry out the present process of evaluation, it was deemed necessary to divide the ST into "semantic chunks" and to judge their translations in terms of fidelity and intelligibility. A semantic chunk is defined as "a sequence of words that fills a semantic role defined in a semantic frame" (Hacioglu & Ward, 2003: p. 1). So, each chunk is supposed to convey some or all of "'who did what to whom?" (Forner & White, 2001). Furthermore, chunks may also correspond to other components related to "where, when, how, and/or why"? In this sequence, the text was divided into 84 semantic chunks (**Appendix 1**). Each online translation package could therefore score a maximum of 84 points for fidelity and the same number of points for intelligibility for each ST produced. To avoid subjective judgments, the Arabic counterpart text of the original ST text was used as the reference in the evaluation process which would judge whether each of these chunks was conveyed into the TT correctly or incorrectly.

Since the present study aimed to investigate how effective the three MT translators targeted are in conveying the meaning of an English ST into Arabic TT, the concern would be focussed more on the semantic quality rather than on the

grammatical quality of the STs produced, unless the grammatical errors obscure the message conveyed from ST to TT. However, it would be essential to indicate that due to the nature of the two languages considered in this study, such errors could lead to fidelity and/or intelligibility problems. This could be clarified in the following two examples. 1) The seventh semantic chunk "to fully fund" in the ST, appears in the same position of the chunk in the UN Arabic copy of the report and in Ginger's TT, but as the ninth chunk at the end of the sentence in both Google and Bing TT outputs (**Appendices 1**, **3** & **6**). This use of the chunk in different positions in the discourse has not obscured fidelity nor intelligibly, and so it was not considered as an error. 2) The semantic chunk (on management activities) that appears in paragraph 2 of the ST, has been translated by the UN concerned department as (fi alanshitah alidariyah—N + adj في ألانشطة الإدارية). However, the three MT packages concerned produced a TT with different structures but all give a similar meaning and effect: (fi anshitat alidarah—N + N -في أنشطة الإدارة). These two examples demonstrate that once grammatical differences do not affect the two translation attributes of fidelity and intelligibility, they would not be treated as real modifications, and so neglected.

On the other hand, some modifications are found to obscure the translation attributes of fidelity as well as the intelligibility of MT. For example, the semantic chunk (for the 2018, 2019) has been translated by the three MT applications (fi assanawat alma'niyah في السنوات المعنية) to mean the aforementioned "years" (plural noun) instead of (fi assanatayn alma'niyah السنتين المعنيتين في) which means the "aforementioned two years". However, the translation packages did not recognize the difference between the dual and plural types of nouns in their Arabic outputs, which resulted in the semantic and grammatical deficiencies in the message conveyed. In a different type of error, the additive article (and: wa-و) in chunk four, has been added to the first sentence after the title of the report in TL outputs of the three MT applications. This addition of (and: wa-و) would disturb the readers of the Arabic texts because they would not be able to find any previous related idea or context to refer or add to. Therefore, this leads to consider this error that doesn't only affect users' understanding of the text, but also the fluency of their reading.

6. Analysis of Machine Translation Outputs

This study aimed to investigate the effectiveness of the three targeted online translators: Google, Ginger, and Bing, and to identify the translator that leads to the least translation errors in terms of the fidelity and intelligibility attributes. To reach this aim, a comparative approach of evaluation was carried out. The ST was divided into 84 chunks (**Appendix 1**), and the TT outputs of each translator were analyzed to identify all errors related to *fidelity* and *intelligibility* of each text.

6.1. Analysis of Google's Translation

The TT as translated via Google Translate seemed to be the least accurate among

the three translations. The translation output displays (**Appendix 3**) some serious lexical, syntactical, and grammatical errors which lead to an extraordinary loss of the message presented in the ST. As summarized in Table 1, the overall percentage of translation errors is 42.85% (percentage of the 15 fidelity errors out of the 84 chunks is 17.85%, and the 21 intelligibility errors out of the 84 chunks is 25.00%). This means that the overall level of accuracy for Google translation is 57.15%. Table 1 also reveals that fidelity problems are caused by different types of errors. Among these problems are the wrong or inaccurate choice of lexical items (e.g., chunks 11, 14, 80, etc.), and wrong words or phrase order (e.g., chunks 21, & 31, 32). Adding or deleting conjunctures has also led to fidelity errors (e.g., chunks 5, 21, 80, etc.). Another remarkable source of errors related to fidelity is the inconsistent and inaccurate use of the concept (period/duration of time = قرتف) in the TT (e.g., chunks 4, 5, 47, and 51). Intelligibility errors, on the other hand, are more in numbers, but less in types. These errors are attributed to the wrong use of singular, dual, and plural nouns (e.g. chunk 11), misuse of feminine and masculine pronouns (e.g., chunks 75). Table 1 also displays that there are also some cases where errors occur due to the lack of fidelity and intelligibility at the same time (e.g., chunks 4, 5, 8. 11, –31, 32–, 47, 57, 64, 74, and 75). These numerous errors resulted in an inaccurate conveyance of the message from the ST into the TT; nevertheless, the TT continues to maintain a level of fluency and intelligibility, especially for sympathetic or patient readers. On the whole, the overall percentage of accuracy for Google is 57.15%, which is obviously less than those of Ginger and Bing, as will be shown in this analysis.

6.2. Analysis of Ginger's Translation

The output of the Ginger translator (**Appendix 5**) looks more accurate than that of the Google translator. Similar to the output of Google, Table 2 shows that there are 15 fidelity errors (17.85% of the 84 chunks) of different types. However, eleven intelligibility errors (13.09% of the 84 chunks) in the Ginger translation are spotted, which remarkably indicates that the Ginger's translated text scored a higher level of fluency compared to that of Google's translation. It is also noticed that the percentage of the observed fidelity errors in the translated texts of Ginger and Google are the same (17.85%). These fidelity errors of Ginger are also of different types, such as 1) inaccurate choice of lexical items (e.g., chunks 11, 40, 50, etc.), 2) word or phrase order (e.g., chunks 31, 32, & 40), 3) adding or deleting conjunctures or demonstrative pronouns (e.g., chunks 2, & 72), iv) inconsistent and inaccurate use of the concept (period/duration of time = قرتف) in the TT (e.g., chunks 4, 5, & 56), and v) missing words in some chunks (e.g., chunks 8, 49, and 70). Again, the intelligibility errors occurring in Ginger's translation are attributed to the wrong use of singular, dual, and plural nouns (e.g., chunks 7, & 11), but there are no errors related to the use of feminine and masculine pronouns. There are also some cases where errors are attributed to the lack of fidelity and intelligibility at the same time (e.g., chunks 4, 5,

Table 1. Error analysis for Google's translation.

	Chunk	Fidelity	Intelligibility	Accepted Translation
4	2021 – 2018 ،.. Missing the word that means period (duration of time)	ü	ü	للفترة 2018-2021
5	وخلال الفترة 2018-2019 A new independent paragraph Preceded by no any context) starts with the additive article (and = و).	ü	ü	خلال الفترة 2018-2021
8	لتمويل النفقات المتكررة The use of *(li* instead of *min ajl*) changes the meaning to look like funding is a determiner for the management efficiencies rather than the result of their management efficiencies work	ü	ü	من أجل تمويل النفقات المتكررة
11	في السنوات المعنية It appears in plural case in TL although it means two years in the SL	ü	ü	في السنتين المعنيتين
12	وسمحت أوجه الكفاءة هذه Inaccurate word: there is a difference in the denotation between the words allowed and permitted.	ü		كما أتاحت أوجه الكفاءة
14	بدعم قرار المدير المتعمد Inaccurate word. The adjective (deliberate) is wrongly translated	ü		بدعم قرار المدير المدروس
17-18	عليها المجلس التنفيذي لاستخدامها في أنشطة الإدارة. The pronoun is repeated		ü	المجلس التنفيذي باستخدامها في أنشطة الإدارة.
21	..يتناقض عدم استخدام مرفق الموارد المعتمد There should be some kind of modifier to let the meaning go straight forward		ü	وكان عدم استخدام مرفق الموارد المعتمد هذا مخالفا (متناقضا)
28	..سمح التخفيض For cohesion purposes, the should be an additive conjuncture		ü	وسمح التخفيض
31 32	إلى 64 في المائة من 62 في المائة Wrong order of the two chunks	ü	ü	من 62 في المائة إلى 64 في المائة
35	لأنشطة الميزانية المؤسسية Inaccurate use of the preposition that leads to a different meaning	ü		من أجل أنشطة الميزانية المؤسسية
47	خلال .. 2018-2019 ، Missing of the word (period = الفترة)	ü	ü	خلال الفترة 2018-2019 ،
50	ارتفاعًا من 88 سنتًا Wrong word		ü	مقارنة بـ— 88 سنتا
51	في 2014-2017. Missing of the word (period = الفترة)		ü	في الفترة 2014-2017.
56	في .. 2018-2019. Missing of the word (period = الفترة)		ü	في الفترة 2018-2019.
57	A weak expressionومقابل مستوى		ü	ومقارنة بمستوى

Continued

57	مخطط مبدئي A past passive structure that requires (past – to be – verb), and inaccurate word (مبدئي)	0	0	كان مخططا له في البداية
64	Lacks the possession pronoun.... من المستوى المخطط	0	0	من المستوى المخطط له
70	خسائر سعر الصرف The expression to tell that the losses in this context are due to the exchange rate is missing	0		الناجمة عن) iiالخسائر المترتبة على سعر الصرف
72	.... تعزى إلى الموارد العادية It needs an additive conjuncture		0	وتعزى إلى الموارد العادية
73	والمبلغ المتبقي .... 0.12 مليار دولار A word that describes the number is needed		0	والمبلغ المتبقي وقدره 0.12 مليار دولار
71	Awkwardness التي تبلغ نحو 0.02 بليون دولا ر		0	والبالغة نحو 0.02 بليون دولار
74	...إلى موارد أخرى. Some missing part causes unclear meaning	0	0	الذي تعزى الى موارد أخرى.
75	إن جائحة COVID-19 له The masculine pronoun is used instead of feminine	0	0	إن جائحة COVID-19 لها
76	Inaccurate Expression تداعيات في الوقت الحقيقي	0		تداعيات آنية
80	It needs an additive conjuncture ..سيخاطب المدي		0	وسيخاطب المدير
	Total	15 =17.85%	21 = 25.00%	36 42.85%

*Fidelity errors out of the total of 84 chunks =17.85%,
*Intelligibility errors out of the total of 84 chunks =25.00%
*Percentage of total translation errors: 36 =42.85%

*Percentage of correct translation: 82.15%
*Percentage of correct translation: 75.00%
*Percentage of total correct translation: 57.15%

8, 11, −31, 31−, 49, 56, 70, & 72). Again, these numerous errors lead to the inaccurate conveyance of the message from the ST into the TT; yet, the TT continues to maintain a better level of accuracy. In general, the overall percentage of Ginger's translation accuracy is 69.06%, which is apparently better than that of Google's.

6.3. Analysis of Microsoft Bing's Translation

In terms of numbers and percentages, analysis of Bing's translation shows different fidelity and intelligibility total of errors when compared with Google and Ginger translations, for the benefit of the Bing translator. **Table 3** displays that 14.28% (12 out of 84 chunks) of the semantic chunks were having fidelity errors that affected the quality of Bing's translation. This table shows that these fidelity and intelligibility errors are almost similar in their linguistic nature to those detected in Google's and Bin's outputs. The fidelity errors in Bing's translation include the inaccurate choice of lexical items (e.g., chunks 4, 40, 50, 70, etc.), absence of words in some chunks (e.g., chunks 8, & 46), incorrect word or phrase order (e.g., chunks 8, & 31, 32), misuse of the concept (period/duration of time = قرتف) in the TT (e.g., chunks 5, 40, & 56), addition or deletion of conjunctures

Table 2. Error analysis for Ginger's translation.

	Chunk	Fidelity	Intelligibility	Accepted Translation
4	**2021 -2018 ....** Missing of the word (period = لفترة)	0	0	للفترة 2018-2021
5	وخلال الفترة 2018-2019 A new independent paragraph (Preceded by no context) starts with the additive article (and = و).	0	0	خلال الفترة 2018-2021
7	من الكفاءة الإدارية A singular noun in TT, but it should be a plural one to comply with the noun in the SL.	0	0	من الكفاءات / أوجه الكفاءة الإدارية
8	لتمويل النفقات المتكررة The use of (*li* instead of min *aj*) changes the meaning to look like funding is a determiner for the management efficiencies rather than the result of their management efficiencies work	0	0	من أجل تمويل النفقات المتكررة
12	من الكفاءة الإدارية The word underlined appears as a singular noun that indicates the efficiency of the administration; it talks about or modifies a process. Whereas the Original sentence talks about the efficacy (skills) of the persons working in the administration		0	من الكفاءات الإدارية
	دعم القرار المتعمد Inaccurate word. The adjective (deliberate) is wrongly translated	0		دعم القرار المدروس
8	لتمويل... النفقات المتكررة The modifier FULLY is missing here. It is used at the end of the sentence, and so it is causing is a kind of ambiguity, not to forget "economy of words".	0		لتمويل كامل النفقات المتكررة
11	في السنوات المعنية It appears in a plural case in TL although it means two years in the SL	0	0	في السنتين المعنيتين
31-32	إلى 64 في المائة من 62 في المائة Wrong order of the two chunks من 62 في المائة and إلى 64 في المائة	0	0	من 62 في المائة إلى 64 في المائة
40	مما ساهم جميعها Inaccurate choice of words, no cohesion	0	0	وساهم كل ذلك
49	مقابل كل دولار من دولارات الولايات المتحدة ..... This sentence has something missing; there is a need to tell the reader about what happens to the US dollars.	0	0	مقابل كل دولار من دولارات الولايات المتحدة من النفقات
50	Wrong word بزيادة عن 88 سنتا	0		مقارنة ب— 88 سنتا
56	في .. 2018-2019. Missing of the word (period = لفترة)	0	0	في الفترة 2018-2019.
68	Inaccurate word سيُقدّم نسبة 98 في المائة	0		سيبلغ نسبة 98 في المائة
70	Inaccurate use of the preposition في أسعار الصرف	0	0	المترتبة على (الناجمة عن) أسعار الصرف
72	... إلى موارد أخرى. Some missing part causes unclear meaning	0	0	التي تعزى الى موارد أخرى.
	Total	15 =16.66%	11 = 13.09%	25 = 29.76%

*Fidelity errors out of the total of 84 chunks =17.85% — 81.95% of Ginger translation is correct
*Intelligibility errors out of the total of 84 chunks =13.09% — 86.91% of Ginger translation is correct
*Total errors and their Percentage: 26 =30.94% — Percentage of total Correct translation: 69.06%

Table 3. Error analysis for Microsoft Bing's translation.

	Chunk	Fidelity	Intelligibility	Accepted Translation
5	وخلال الفترة 2018-2019 A new independent paragraph (Preceded by no context) starts with the additive article (and = و).	ü	ü	خلال الفترة 2018-2021
7	من الكفاءة الإدارية A singular noun in T, but it should be a plural one to comply with the noun in the SL.	ü		من الكفاءات / أوجه الكفاءة الإدارية
8	لتمويل النفقات المتكررة The use of (*li* instead of min *aj*) changes the meaning to look like funding is a determiner for the management efficiencies rather than the result of their management efficiencies work	ü	ü	من أجل تمويل النفقات المتكررة
8	لتمويل ..... النفقات المتكررة The modifier FULLY is missing here. It is used at the end of the sentence, and so it is causing is a kind of ambiguity, not to forget "economy of words".	ü		من أجل تمويل كامل النفقات المتكررة
11	في السنوات المعنية It appears in a plural case in TL although it means two years in the SL	ü	ü	في السنتين المعنيتين
14	بدعم قرار المدير المتعمد Inaccurate word. The adjective (deliberate) is wrongly translated	ü		بدعم قرار المدير المدروس
31 32	إلى 64 في المائة من 62 في المائة Wrong order of the two chunks من 62 في المائة and إلى 64 في المائة	ü	ü	من 62 في المائة إلى 64 في المائة
40	مما ساهم جميعها Inaccurate choice of words, no cohesion	ü	ü	وساهم كل ذلك
46	مقابل كل دولار من دولارات الولايات المتحدة ..... This sentence has something missing; there is a need to tell the reader about what happens to the US dollars.	ü	ü	مقابل كل دولار من دولارات الولايات المتحدة من النفقات
50	Inaccurate word بزيادة عن 88 سنتا	ü		مقارنة ب— 88 سنتا
56	في .. 2018-2019. Missing of the word (period = الفترة)		ü	في الفترة 2018-2019.
70	في أسعار الصرف Inaccurate use of the preposition	ü	ü	المترتبة على (الناجمة عن) أسعار الصرف
71	التي تبلغ نحو 0.02 بليون دولار Awkwardness		ü	والبالغة نحو 0.02 بليون دولار
74	..إلى موارد أخرى. Some missing part causes unclear meaning	ü	ü	التي تعزى الى موارد أخرى.
	Total	12 =14.28%	10 =11.90%	22 =26.19%

*Fidelity errors out of the total of 84 chunks =14.28% — 85.72% of Bing translation is correct
*Intelligibility errors out of the total of 84 chunks =11.90% — 88.10% of Bing translation is correct
*Total errors and their Percentage: 22 =26.19% — Percentage of total Correct translation: 73.81%

(e.g., chunks 5, & 40). Concerning the fluency feature of Bing's output, the percentage of errors appears less than that of Ginger's output, and remarkably less than that of Google's output. These errors are attributed to incorrect use of singular, dual, and plural nouns (e.g., chunks 7, & 11), and incorrect word/phrase order (e.g., chunks 8, & 31, 32). It is also noticed, as the case is with the other two translators, that the translation of eight chunks distorted both fidelity and intelligibility of the translated text. That is, eight chunks of Bing's output (chunks 5, 8, 11, −31, 32−, 40, 46, 70, & 74) are identified to have both accuracy and fluency deficiencies at the same time. Despite these fidelity and intelligibility errors, the quality of Bing's output appears more accurate and fluent than that of Google's and Ginger's outputs. This could be identified easily by comparing the overall level of quality of Bing's translation (73.81%) with the same percentages of the quality related to the translations of Ginger and Google (69.06% and 57.15%), respectively, as displayed in **Tables 1-3**. Yet, readers' understanding of the translated text continues to face some hindrances due to the 26.19% percentage of adequacy and fluency errors.

7. Discussion

In the attempt to evaluate the performance of the three free online MT tools (Google Translator, Bing, and Ginger) in translating an English into Arabic text, the present study quantitatively analyzed the total number and the percentage of total translation errors in the target language texts (TLTs), after the qualitative semantic and syntactic study of the source and the target texts. While the findings showed that the three TLTs were having serious fidelity and intelligibility errors that affected the adequacy and fluency of the three texts, differences in the percentages of errors were easily identified between them, and mainly between Google's translation (42.85% percentage of errors) on the one hand and those of Ginger and Bing on the other hand (30.94%, and 26.19% percentage of errors, respectively). This surprising percentage of errors for Google's translation which is clearly exceeding those of Bing's and Ginger's translations was not expected; at least because most of the Arab users (mainly the researcher's students) attend to use Google Translate perhaps because it is an application that is available by default on the Google browser. This result, therefore, proves that the translation of the Microsoft Bing application scored a higher quality (73.81% of its translation is correct) of translation than those of the other two applications in terms of fidelity and intelligibility. Although the percentage of correct translation for Ginger was less than that of Bing's, a slight difference for the benefit of Bing was revealed between them, (69.06% and 73.81%, respectively).

The findings of the present study came in the same sequence with the studies of Niño (2009), Belam (2003), Kliffer (2005), and Bowker (2014) which investigated the effectiveness of using the post-editing technique of raw MT outputs as a classroom strategy to help foreign language students enhance their language and translation abilities. This study clearly showed that MT raw outputs have

different linguistic errors, and these could be used for language learning and translation training purposes.

The importance of human intervention to improve MT outputs has also been confirmed by Al-Khresheh and Almaaytah (2018) who investigated the linguistic obstacles that MT might face in translating English proverbs into Arabic. Findings of this study are found consistent with those findings of Al-Khresheh and Almaaytah's findings which revealed that the translation of English into Arabic proverbs using Google Translate lacked accuracy and logic due to the awkwardness of word meanings produced.

However, different findings were reported by Taleghani, and Pazouki (2018), who studied the effectiveness of using Google Translate and Bing in addition to the other two applications in translating English idioms and phrasal verbs into Persian. The study reported that the translation of the Iranian MT "Targoman Program" scored a satisfactory quality of translation compared to the other of the three applications. This finding could be attributed to some features of the software that mainly serves the Persian language. However, Taleghani, and Pazouki (2018) revealed that the translation quality of the other applications was relatively of low quality and this was found consistent with what has been achieved in the present study. In the same sequence findings of the present study were found inconsistent with Daniele's (2019) study which reported that the mean percentage of errors in Google's translation of medical texts (abstracts of research) from English into Italian was 15%. This fairly good performance of MT could be referred to the fact that English and Italian belong to the same family of languages that share to some extent many cultural aspects, in addition to some areas in their linguistic systems. However, in the case of the present study, Arabic and English belong to different language families with different language systems, a case that contributed to the surprising percentages of total translation errors that reached 42.85% in Google's translation.

It might be vital to clarify that some factors could have caused the different findings of the present study. This study implemented some common three MT applications in the Arab world, mainly in Saudi Arabia, but none of them is known to be designed in a way to give any privilege to the Arabic language. The present study also used the MT applications to translate a whole lengthy common type of English text that addressed related ideas or arguments presented in one complete text. Choosing to translate this type of lengthy text where most of the English language structures can be met, was expected to confuse the work of the MT system in the attempts to coordinate individual elements of the source language text with the TL at the different correspondence levels (Wilss, 1982). However, this job looks totally easier for the MT tools to coordinate individual elements in the studies that were conducted to investigate MT effectiveness at the level of separate expressions, idioms, and phrasal verbs. These factors must have caused the inconsistent findings of the present and the previous studies reviewed in the literature section.

8. Conclusion

This study aimed to investigate the quality of free MT systems in translating English into Arabic texts. In this sequence, it was an attempt to evaluate and compare the English into Arabic translation outputs of Google Translate, Microsoft Bing, and Ginger. For this purpose, a lengthy text was chosen from the UN records, and segmented into 84 semantic chunks to reduce the subjective assessment of the quality of the outputs; and the assessment of each semantic chunk had two levels: correct or incorrect accuracy of the translation (fidelity), and correct or incorrect fluency of the translation (intelligibility).

The descriptive quantitative analysis of the TTs revealed that serious accuracy and fluency errors were met in the three outputs. Those errors were of different linguist types at the level of word choice and use, and use of prepositions, word order in the chunk, singular, dual, and plural use of nouns/pronouns, and the use of conjunctures. Surprisingly, the mean percentage of errors in Google's translation was the highest (42.85%), followed by that of Ginger's translation (30.94%), and the least percentage of errors detected was for Microsoft Bing (26.19%). In other words, these results indicated that the translation of the Microsoft Bing system scored relatively a higher quality than those of the other two systems. However, the study of these percentages of errors led to the conclusion that the Arabic translation of the three MT applications was not of high quality. Furthermore, the study concludes that the less the post-editing is required, the more successful the translation is, the less the time is spent and the less the effort is spent to produce a translation of high quality.

As a result of this description of the errors and their types, this study recommends that language and translation instructors may use similar Arabic translated outputs and their English origins in classroom activities to help students improve their language and translation skills. It is also important to keep in mind that through the continuous development in the field of artificial intelligence, MT is always witnessing improvements in its applications, so further research is recommended to evaluate more MT applications as well as different types of texts. Finally, and since the Arabic language suffers from fewer materials being uploaded on the Internet, compared to its large number of users (Ali, 2016), machine translation applications do not have sufficient ample opportunities for adequate self-learning through TMs. This situation requires that Arabic language users should increase their online Arabic language contributions in terms of quantity and genre.

Acknowledgements

This Publication was supported by the Deanship of Scientific Research at Prince Sattam bin Abdulaziz University.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

Alawneh, M. F., & Sembok, T. M. (2011). Rule-Based and Example-Based Machine Translation from English to Arabic. In *2011 Sixth International Conference on Bio-Inspired Computing: Theories and Applications* (pp. 343-347).

Albat, T. F. (2012). *Systems and Methods for Automatically Estimating a Translation Time*. US Patent 0185235, 19 July 2012.

Ali, M. A. (2016). Artificial Intelligence and Natural Language Processing: The Arabic Corpora in Online Translation Software. *International Journal of Advanced and Applied Science, 3,* 59-66. https://doi.org/10.21833/ijaas.2016.09.010 http://science-gate.com/IJAAS/Articles/2016-3-9/10%202016-3-9-pp.59-66.pdf

Ali, M. A. (2018). The Human Intelligence vs. Artificial Intelligence: Issues and Challenges in Computer Assisted Language Learning. *International Journal of English Linguistics, 8,* 259-271. http://www.ccsenet.org/journal/index.php/ijel/article/view/75235 https://doi.org/10.5539/ijel.v8n5p259

Al-Khresheh, M. (2016). A Review Study of Contrastive Analysis Theory. *Journal of Advances in Humanities and Social Sciences, 2,* 330-338. https://doi.org/10.20474/jahss-2.6.5

Al-Khresheh, M. H., & Almaaytah, S. A. (2018). English Proverbs into Arabic through Machine Translation. *International Journal of Applied Linguistics & English Literature, 7,* 158-166. https://doi.org/10.7575/aiac.ijalel.v.7n.5p.158

Anderson, D. D. (1995). Machine Translation as a Tool in Second Language Learning. *CALICO Journal, 13,* 68-96.

Belam, J. (2003). Teaching Machine Translation Evaluation by Assessed Project Work. In *6th EAMT Workshop Teaching Machine Translation* (pp. 131-136). Manchester.

Bowker, L. (2014). *Can Machine Translation Facilitate Outreach to Newcomers? A Pilot Study Investigating the Needs of Spanish-Speaking Users of the Ottawa Public Library*. 2013 OCLC/ALISE Research Grant Report Published Electronically by OCLC Research. http://www.oclc.org/research/grants/reports/2013/bowker2013.pdf

Cieślak, M. (2011). The Scope and Limits of Machine Translation. *eLingUp [Centro de Linguística da Universidade do Porto], 3,* 156-174. http://193.137.34.194/index.php/elingUP/article/download/2528/2316

Daniele, F. (2019). Performance of an Automatic Translator in Translating Medical Abstracts. *Heliyon, 5,* e02687. https://doi.org/10.1016/j.heliyon.2019.e02687 https://www.researchgate.net/publication/336706218_Performance_of_an_automatic_translator_in_translating_medical_abstracts

Doherty, S. (2016). The Impact of Translation Technologies on the Process and Product of Translation. *International Journal of Communication, 10,* 947-969. https://www.researchgate.net/publication/284725157_The_impact_of_translation_technologies_on_the_process_and_product_of_translation

Fiederer, R., & O'Brien, S. (2009). Quality and Machine Translation: A Realistic Objective? *The Journal of Specialised Translation, 11,* 52-74.

Forner, M., & White, J. S. (2001). Predicting MT Fidelity from Noun-Compound Handling. In *Workshop on MT Evaluation, "Who Did What to Whom?"*. Mt Summit VIII, Santiago de Compostela, Spain.

Hacioglu, K., & Wayne, W. (2003). Target Word Detection: Semantic Role Chunking Using Support Vector Machines. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language*

Technology: Companion Volume of the Proceedings of HLT-NAACL 2003—Short Papers (Vol. 2, pp. 25-27). HIL-NAACL.
http://dl.acm.org/citation.cfm?id=1073483.1073492
https://doi.org/10.3115/1073483.1073492

Han, A. L. F., Wong, D. F., & Chao, L. S. (2012). LEPOR: A Robust Evaluation Metric for Machine Translation with Augmented Factors. In *Proceedings of the 24th International Conference on Computational Linguistics (COLING 2012)* (pp. 441-450). Posters, Mumbai, India. https://github.com/aaronlifenghan/aaron-project-lepor

Hirschman, L., Reeder, F., Burger, J., & Miller, K. (2000). Name Translation as a Machine Translation Evaluation Task. In *Proceedings of the Workshop on Machine Translation Evaluation* (pp. 21-28). New Jersey: LREC-2000.
https://www.jostrans.org/issue11/art_fiederer_obrien.php

Hutchins, J., & Somers, H. (1992). *An Introduction to Machine Translation*. London: Academic Press Limited.

Kliffer, M. D. (2005). An Experiment in MT Post-Editing by a Class of Intermediate/Advanced French Majors. In *Proceedings EAMT 10th Annual Conference* (pp. 160-165). Budapest.

La Torre, M. D. (1999). A Web-Based Resource to Improve Translation Skills. *ReCALL, 11,* 41-49.

Lee, S.-M. (2019). The Impact of Using Machine Translation on EFL Students' Writing. *Computer Assisted Language Learning, 33,* 157-175.
https://doi.org/10.1080/09588221.2018.1553186

Niño, A. (2004). *Recycling MT: A Course on FL Writing via MT Post-Editing* (pp. 179-187). Paper Presented at CLUK (Computational Linguistics United Kingdom 7th Annual Research Colloquium), 6th and 7th January 2004 in the University of Birmingham, UK.
https://scholar.google.com/scholar?q=Ni%C3%B1o+A+.+(2004)+Recycling+MT:+A+course+on+FL+writing+via+MT+post-editing.+Paper+presented+at+CLUK+(Computational+Linguistics+United+Kingdom+7th+Annual+Research+Colloquium)+6th+and+7th+January+2004+in+the+University+of+Birmingham+UK

Niño, A. (2009). Machine Translation in Foreign Language Learning: Language Learners' and Tutors' Perceptions of Its Advantages and Disadvantages. *ReCALL, 21,* 105-122.
https://doi.org/10.1017/S0958344009000172

Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. (2002). BLEU: A Method for Automatic Evaluation of Machine Translation. In *ACL-2002: 40th Annual Meeting of the Association for Computational Linguistics* (pp. 311-318). Stroudsburg,
https://doi.org/10.3115/1073083.1073135

Taleghani, M., & Pazouki, E. (2018). Free Online Translators: A Comparative Assessment in Terms of Idioms and Phrasal Verbs. *International Journal of English Language & Translation Studies, 6,* 15-19.

Turian, J., Shen, L., & Melamed, I. D. (2003). Evaluation of Machine Translation and Its Evaluation. In *Proceedings of the MT Summit IX* (pp. 386-393).

White, J. (1995). Approaches to Black-Box Machine Translation Evaluation. In *Proceedings of MT Summit 1995* (pp. 386-393). New Orleans, USA.

White, J. (2001). Predicting Intelligibility from Fidelity. In *Proceedings of the Workshop on Evaluation*. MT Summit VI, Santiago, Spain.
http://www.eamt.org/events/summitVIII/papers/white-1.pdf

White, J. S. (2003). How to Evaluate Machine Translation. In H. Somers (Ed.), *Comput-*

ers and Translation (pp. 211-244). Amsterdam: John Benjamins. https://doi.org/10.1075/btl.35.16whi

Wilss, W. (1982). *The Science of Translation. Problems and Methods*. Vol. 80, Tubingen: Narr.

Zong, Z. (2018). Research on the Relations between Machine Translation and Human Translation. *Journal of Physics: Conference Series, 1087,* Article ID: 062046. https://doi.org/10.1088/1742-6596/1087/6/062046 https://iopscience.iop.org/article/10.1088/1742-6596/1087/6/062046/pdf

Appendix 1. Semantic Chunks for the Source Text

No.	Chunk	No.	Chunk
1	Midterm review	43	as articulated in the integrated results
2	of the integrated resources plan	44	and resources framework
3	and integrated budget	45	for 2018 and 2019
4	2018-2021	46	For every United States dollar of expenditure
5	During the period 2018-2019	47	during 2018-2019,
6	UNDP leverzged	48	91 cents were spent
7	management efficiencies	49	on development programmes and services
8	to fully fund	50	up from 88 cents
9	the recurring institutional budget expenditures	51	in 2014-2017.
10	from the institutional budget revenue earned	52	In aggregate
11	in the respective years	53	this represents
12	These efficiencies also allowed	54	approximately $240 million
13	UNDP	55	in additional resources available for develapment
14	to support the Administrator's deliberate decision	56	in 2018-2019
15	not to spend the $49 million	57	Against an initially planned level
16	of additional regular resources	58	of $15.3 billion
17	approved by the Executive Board	59	in available resources
18	for use	60	for 2018-2019
19	on management activities	61	UNDP attained $14.9 billon
20	The non-utilization of this approved resource facility	62	in available resources
21	contrasted with past strategic pan periods	63	reaching 97.4 percent
22	when recurring institutional budget expenditures	64	of the planned level
23	were funded	65	indicating
24	from institutional budget reverue	66	the continued confidence of funding partners
25	earned in the year	67	UNDP would have reached
26	and the reserves were used	68	98 percent
27	at the discretion of the Administrator	69	of the planned level
28	The reduction in institutional budget expenditure	70	if not for the impact of exchange rate losses
29	allowed the proportion of regular resources	71	of approximately $0.02 billion
30	allocated to development programmes to Increase	72	attributed to regular resources
31	to 64 percert	73	and the remaining $0.12 billion
32	from 62 percert	74	to other resources
33	(approximately $20 million),	75	The COVID-19 pandemic
34	while the proportion of regular resources	76	is having real-time repercussions
35	used for institutional budget activties	77	for the work of UNDP
36	decreased	78	The Administrator will address
37	from 38 percent	79	the Executive Board
38	to 36 percent	80	separately
39	compared to the previous period	81	on the pandemic
40	all of which contnibuted	82	the implications
41	to the development results	83	for the institutional component of the integrated budget
42	achieved by UNDP	84	and its impact on the Strategic Plan

Appendix 2. English Text (ST)

Annual session 2020, 1-5 June 2020, New York, Item 6 of the provisional agenda

Midterm review of the Integrated resources plan and Integrated budget, 2018-2021

During the period 2018-2019, UNDP leveraged management efficiencies to fully fund the recurring institutional budget expenditures from the institutional budget revenue earned in the respective years.

These efficiencies also allowed UNDP to support the Administrator's deliberate decision not to spend the $49 million of additional regular resources approved by the Executive Board for use on management activities. The non-utilization of this approved resource facility contrasted with past strategic plan periods when recurring institutional budget expenditures were funded from Institutional bud get revenue earned in the year and the reserves were used at the discretion of the Administrator.

The reduction in institutional budget expenditure allowed the proportion of regular resources allocated to development programmes to increase to 64 per cent from 62 per cent (approximately $20 million), while the proportion of regular resources used for institutional budget activities decreased from 38 per cent to 36 per cent compared to the previous period, all of which contributed to the development results achieved by UNDP as articulated in the integrated results and resources framework for 2018 and 2019.

For every United States dollar of expenditure during 2018-2019, 91 cents were spent on development programmes and services, up from 88 cents in 2014-2017. In aggregate, this represents approximately $240 million in additional resources available for development in 2018-2019.

Against an initially planned level of $15.3 billion in available resources for 2018-2019, UNDP attained $14.9 billion in available resources, reaching 97.4 per cent of the planned level, indicating the continued confidence of funding partners. UNDp would have reached 98 per cent of the planned level if not for the impact of exchange rate losses of approximately $0.02 billion attributed to regular resources and the remaining $0.12 billion to other resources.

The COVID-19 pandemic is having real-time repercussions for the work of UNDP, The Administrator will address the Executive Board separately on the pandemic, the implications for the institutional component of the integrated budget and its impact on the Strategic Plan.

Reached at:

https://digitallibrary.un.org/record/3861578/files/DP_2020_9-EN.pdf

Appendix 3. UN Arabic Text: Formal Counterpart Text (TT)

استعراض منتصف المدة للخطة المتكاملة للموارد والميزانية المتكاملة للفترة 2018- 2021

خلال الفترة 2018-2019 ، استفاد البرنامج الإنمائي من أوجه الكفاءة الإدارية من أجل التمويل الكامل لنفقات الميزانية المؤسسية المتكررة من إيرادات الميزانية المؤسسية المكتسبة في السنتين المعنيتين.

وأتاحت أوجه الكفاءة هذه أيضا للبرنامج الإنمائي دعم القرار المتعمد الذي اتخذه مدير البرنامج بعدم إنفاق مبلغ 49 مليون دولار من الموارد العاديمة الإضافية الذي وافق عليه المجلس التنفيذي لإستخدامه فيما يتعلق باللأنشطة الإدارية. وكان عدم استخدام مرفق الموارد المعتمد هذا مخالفا لفترات الخطة الإستراتيجية الماضية عندما تم تمويل النفقات المتكررة للميزانيمة المؤسسية من إيرادات الميزانية المؤسسية المكتسبة في السنة واستخدمت الإحتياطات حسب السلطة التقديرية لمدير البرنامج.

وسمح الإنخفاض في نفقات الميزانية المؤسسية بارتفاع نسبة الموارد العادية المخصصة لبرنامج التنمية إلى 64 في المائة من 62 في المائة (حوالي 20 مليون دولار)، فيما انخفضت نسبة الموارد العادية المستخدمة لأنشطة الميزانية المؤسسية من 38 في المائة إلى 36 في المائة مقارنة بالفترة السابقة، وساهم كل ذلك في نتائج التنمية التي حققها البرنامج الإنمائي على النحو المبين في الإطار المتكامل للنتائج والموارد لعامي 2018 و 2019.

وبالنسبة لكل دولار من دولارات الولايات المتحدة من النفقات خلال الفترة 2018-2019 ، أنفق 91 سنتا على البرنامج والخدمات الإنمائية، وهو ما يمثل زيادة مقارنة بـ 88 سنتا في الفترة 2014-2017 . وإجمالا، يمثل ذلك حوالي 240 مليون دولار من الموارد الإضافية المتاحة للتنمية في الفترة 2018-2019.

ومقارنة بمستوى كان مقررا في البداية هو 15,3 بليون دولار من الموارد المتاحة للفترة 2018-2019 ، حقق البرنامج الإنمائي 14,9 بليون دولار من الموارد المتاحة، ليصل إلى 97,4 في المائة من المستوى المقرر، مما يشير إلى استمرار ثقة شركاء التمويل. وكان البرنامج الإنمائي سيبلغ 98 في المائة من المستوى المقرر لولا أثر الخسائر المتعلقة بأسعار الصرف البالغة حوالك 0,02 بليون دولار التي تعزى إلى الموارد العادية والمبلغ المتبقي وقدره 0,12 بليون دولار الذي يعزى إلى الموارد الأخرى .

وتترتب على جائحة كوفيد-19 آثار في الوقت الحقيقي على عمل البرنامج الإنمائي. وسيطلع مدير البرنامج المجلس التنفيذي بشكل منفصل على معلومات بشأن الجائحة، والآثار المترتبة على العنصر المؤسسي للميزانية المتكاملة وأثرها على الخطة الإستراتيجية.

Reached at:

https://digitallibrary.un.org/record/3861578/files/DP_2020_9-AR.pdf

Appendix 4. Google's Arabic Translation (TT)

مراجعة منتصف المدة لخطة الموارد المتكاملة والميزانية المتكاملة 2018-2021

وخلال الفترة 2018-2019، استفاد برنامج الأمم المتحدة الإنمائي من الكفاءات الإدارية لتمويل نفقات الميزانية المؤسسية المتكررة بالكامل من إيرادات الميزانية المؤسسية المكتسبة في السنوات المعنية.

وسمحت أوجه الكفاءة هذه أيضا للبرنامج الإنمائي بدعم قرار المدير المتعمد بعدم إنفاق 49 مليون دولار من الموارد العادية الإضافية التي وافق عليها المجلس التنفيذي لاستخدامها في أنشطة الإدارة. ويتناقض عدم استخدام مرفق الموارد المعتمد هذا مع فترات الخطة الاستراتيجية السابقة عندما تم تمويل نفقات الميزانية المؤسسية المتكررة من إيرادات الميزانية المؤسسية المكتسبة في السنة واستخدمت الاحتياطيات وفقًا لتقدير المدير.

سمح التخفيض في نفقات الميزانية المؤسسية بزيادة نسبة الموارد العادية المخصصة لبرامج التنمية إلى 64 في المائة من 62 في المائة (حوالي 20 مليون دولار) ، في حين انخفضت نسبة الموارد العادية المستخدمة لأنشطة الميزانية المؤسسية من 38 في المائة إلى 36 في المائة مقارنة بالفترة السابقة ، والتي ساهمت جميعها في نتائج التنمية التي حققها برنامج الأمم المتحدة الإنمائي كما هو موضح في إطار النتائج والموارد المتكامل لعامي 2018 و 2019.

مقابل كل دولار أمريكي من الإنفاق خلال .. 2018-2019 ، تم إنفاق 91 سنتًا على برامج وخدمات التطوير ، ارتفاعًا من 88 سنتًا في 2014-2017. بشكل إجمالي ، يمثل هذا حوالي 240 مليون دولار من الموارد الإضافية المتاحة للتطوير في 2018-2019.

ومقارنة بمستوى كان مقررا في البداية وهو 15.3 بليون دولار من الموارد المتاحة للفترة 2018-2019، حقق البرنامج الإنمائي 14.9 بليون دولار من الموارد المتاحة، أيسل إلى 97.4 في المائة من المستوى المقرر، مما يشير إلى استمرار ثقة شركاء التمويل. وكان البرنامج الإنمائي سيبلغ 98 في المائة من المستوى المقرر لولا أثر الخسائر المتعلقة بأسعار الصرف البالغة حوالي 0.02 بليون دولار التي تعزى إلى الموارد العادية والمبلغ المتبقي وقدره 0.12 بليون دولار الذي يعزى إلى الموارد الأخرى.
23 chunks

ومقابل مستوى مخطط مبدئي قدره 15.3 مليار دولار من الموارد المتاحة للفترة 2018 -2019 ، حقق برنامج الأمم المتحدة الإنمائي 14.9 مليار دولار من الموارد المتاحة ، أيسل إلى 97.4 في المائة من المستوى المخطط له ، مما يشير إلى استمرار ثقة شركاء التمويل. وكان برنامج الأمم المتحدة الإنمائي سيصل إلى 98 في المائة من المستوى المخطط لولا تأثير خسائر سعر الصرف التي تبلغ حوالي 0.02 مليار دولار تعزى إلى الموارد العادية والمبلغ المتبقي 0.12 مليار دولار ...إلى الموارد الأخرى.

إن جائحة COVID-19 له تداعيات في الوقت الحقيقي على عمل برنامج الأمم المتحدة الإنمائي. ..سيخاطب المدير المجلس التنفيذي على حدة بشأن الوباء ، والآثار المترتبة على العنصر المؤسسي للميزانية المتكاملة وتأثيره على الخطة الاستراتيجية.

Appendix 5. Ginger's Arabic Translation Text (TT)

استعراض منتصف المدة لخطة الموارد المتكاملة والميزانية المتكاملة، 2018-2021

وخلال الفترة الإنمائي من الكفاءة الإدارية لتمويل النفقات المتكررة للميزانية المؤسسية من إيرادات 2018-2019، استفاد البرنامج الميزانية المؤسسية المكتسبة في السنوات المعنية تمويلا كاملا.
كما أتاحت هذه الكفاءات للبرنامج الإنمائي دعم قرار مدير البرنامج المتعمد بعدم إنفاق مبلغ ?? مليون دولار من الموارد العادية الإضافية التي وافق المجلس التنفيذي على استخدامها في أنشطة الإدارة. ويتناقض عدم استخدام هذا المرفق المعتمد للمورد مع الفترات السابقة للخطة الاستراتيجية عندما كانت نفقات الميزانية المؤسسية المتكررة تمول من إيرادات الميزانية المؤسسية المكتسبة في السنة، واستخدمت الاحتياطيات حسب تقدير مدير البرنامج.
وسمح الانخفاض في نفقات الميزانية المؤسسية بزيادة نسبة الموارد العادية المخصصة للبرامج الإنمائية إلى 64 في المائة من 62 في المائة (حوالي 20 مليون دولار)، في حين انخفضت نسبة الموارد العادية المستخدمة في أنشطة الميزانية المؤسسية من 38 في المائة إلى 36 في المائة مقارنة بالفترة السابقة، مما ساهم جميعها في تحقيق النتائج الإنمائية التي حققها البرنامج الإنمائي على النحو المبين في إطار النتائج والموارد المتكاملة لعامي 2018 و 2019.
مقابل كل دولار من دولارات الولايات المتحدة خلال الفترة 2018-2019، تم إنفاق 91 سنتا على برامج وخدمات التنمية، بزيادة عن 88 سنتا في الفترة 2014-2017. وإجمالاً، يمثل هذا المبلغ حوالي 240 مليون دولار من الموارد الإضافية المتاحة للتنمية في 2018-2019.
ومقارنة بمستوى كان مقررا في البداية وهو 15.3 بليون دولار من الموارد المتاحة للفترة 2018-2019، حقق البرنامج الإنمائي 14.9 بليون دولار من الموارد المتاحة، أيسل إلى 97.4 في المائة من المستوى المقرر، مما يشير إلى استمرار ثقة شركاء التمويل. وكان البرنامج الإنمائي سيبلغ 98 في المائة من المستوى المقرر لولا أثر الخسائر المتعلقة بأسعار الصرف البالغة حوالي 0.02 بليون دولار التي تعزى إلى الموارد العادية والمبلغ المتبقي وقدره 0.12 بليون دولار الذي يعزى إلى الموارد الأخرى.

وفي مقابل مستوى كان مقرراً في البداية وقدره 15.3 بليون دولار في الموارد المتاحة للفترة 2018-2019، حقق البرنامج الإنمائي 14.9 بليون دولار من الموارد المتاحة، أيسل إلى 97.4 في المائة من المستوى المقرر، مما يشير إلى استمرار ثقة شركاء التمويل. وكان البرنامج الإنمائي سيُحقّم نسبة 98 في المائة من المستوى المقرر لولا أثر الخسائر في أسعار الصرف التي تبلغ نحو 0.02 بليون دولار تعزى إلى الموارد العادية والـ 0.12 بليون دولار المتبقية إلى موارد أخرى.

إن وباء "كوفيد-19" له انعكاسات آنية على عمل برنامج الأمم المتحدة الإنمائي. وسيخاطب مدير البرنامج المجلس التنفيذي بصورة منفصلة بشأن الوباء، والآثار المترتبة على العنصر المؤسسي للميزانية المتكاملة وأثره على الخطة الاستراتيجية.

Appendix 6. Microsoft Bing's Arabic Translation Text (TT)

استعراض منتصف المدة لخطه الموارد المتكامله والميزانيه المتكامله، 2018-2021

وخلال الفترة 2018-2019، استفاد البرنامج الإنمائي من الكفاءة الإدارية لتمويل النفقات المتكررة للميزانية المؤسسية من إيرادات الميزانية المؤسسية المكتسبة في السنوات المعنية تمويلا كاملا.

كما أتاحت هذه الكفاءات للبرنامج الإنمائي ???دعم قرار مدير البرنامج المتعمد بعدم إنفاق مبلغ ?? مليون دولار من الموارد العادية الإضافية التي وافق المجلس التنفيذي على استخدامها في أنشطة الإدارة. ويتناقض عدم استخدام هذا المرفق المعتمد للموارد مع الفترات السابقة للخطة الاستراتيجية عندما كانت نفقات الميزانية المؤسسية المتكررة تمول من إيرادات الميزانية المؤسسية المكتسبة في السنة، واستخدمت الاحتياطيات حسب تقدير مدير البرنامج.

وسمح الانخفاض في نفقات الميزانية المؤسسية بزيادة نسبة الموارد العادية المخصصة للبرامج الإنمائية إلى 64 في المائة من 62 في المائة (حوالي 20 مليون دولار)، في حين انخفضت نسبة الموارد العادية المستخدمة في أنشطة الميزانية المؤسسية من 38 في المائة إلى 36 في المائة مقارنة بالفترة السابقة، مما ساهم جميعها في تحقيق النتائج الإنمائية التي حققها البرنامج الإنمائي على النحو المبين في إطار النتائج والموارد المتكاملة لعامي 2018 و 2019.

مقابل كل دولار من دولارات الولايات المتحدة... خلال الفترة 2018-2019، تم إنفاق 91 سنتا على برامج وخدمات التنمية، بزيادة عن 88 سنتا في الفترة 2014-2017. وإجمالاً، يمثل هذا المبلغ حوالي 240 مليون دولار من الموارد الإضافية المتاحة للتنمية في 2018-2019.

ومقارنة بمستوى كان مقررا في البداية وهو 15.3 بليون دولار من الموارد المتاحة للفترة 2018-2019، حقق البرنامج الإنمائي 14.9 بليون دولار من الموارد المتاحة، أيسل إلى 97.4 في المائة من المستوى المقرر، مما يشير إلى استمرار ثقة شركاء التمويل. وكان البرنامج الإنمائي سيبلغ 98 في المائة من المستوى المقرر لولا أثر الخسائر المتعلقة بأسعار الصرف البالغة حوالي 0.02 بليون دولار التي تعزى إلى الموارد العادية والمبلغ المتبقي وقدره 0.12 بليون دولار الذي يعزى إلى الموارد الأخرى.

وفي مقابل مستوى كان مقرراً في البداية وقدره 15.3 بليون دولار في الموارد المتاحة للفترة 2018-2019، حقق البرنامج الإنمائي 14.9 بليون دولار من الموارد المتاحة، أيسل إلى 97.4 في المائة من المستوى المقرر، مما يشير إلى استمرار ثقة شركاء التمويل. وكان البرنامج الإنمائي سيُحقّم نسبة 98 في المائة من المستوى المقرر لولا أثر الخسائر في أسعار الصرف التي تبلغ نحو 0.02 بليون دولار تعزى إلى الموارد العادية والـ 0.12 بليون دولار المتبقية... إلى موارد أخرى.

إن وباء "كوفيد-19" له انعكاسات آنية على عمل برنامج الأمم المتحدة الإنمائي. وسيخاطب مدير البرنامج المجلس التنفيذي بصورة منفصلة بشأن الوباء، والآثار المترتبة على العنصر المؤسسي للميزانية المتكاملة وأثره على الخطة الاستراتيجية.