Research Journal in Advanced Humanities

https://doi.org/10.58256/9bvdfp31

RJAH

Open Access

Check for updates

REVIEW ARTICLE

Section: *Literature, Linguistics & Criticism*

Tracking improvement from statistical to neural machine translation: An error-based evaluation of google translate

Linda S. Al-Abbas[1], Bilal Khalid Khalaf[2], Omar Alali[1] & Marwan Alqaryouti[1]

[1]Department of English Language, Literature and Translation, Zarqa University, Jordan

[2]Department of Translation, College of Arts, University of Anbar, Iraq

*Correspondence: labbas@zu.edu.jo

ABSTRACT

This study investigated the quality of Google Translate outputs after the system shifted from statistical machine translation (SMT) to neural machine translation (NMT). The examples originally analyzed by Al-Zebary (2012), when Google Translate operated under the SMT framework, were retranslated through Google Translate in 2023 and 2025. The resulting outputs were examined for both lexical and structural issues, and a comparative analysis was conducted across the three sets of translations to determine whether the problems previously identified continue to persist. The findings revealed that Google Translate has undergone substantial improvement under NMT, where many of the errors reported in earlier SMT output such as deletion, addition, misinterpretation of homographs, and literal translation of idioms have been significantly reduced. There was also a noticeable improvement from the 2023 NMT outputs to those produced in 2025, indicating continued refinement in Google Translate's handling of English–Arabic translation. However, the study shows that unnatural phrasing continues to occur in linguistically complex contexts, especially those involving cross-linguistic structural mismatches. The results confirm that despite notable progress, machine translation still requires human intervention to ensure accuracy, naturalness, and contextual appropriateness in English–Arabic translation.

KEYWORDS: error analysis, google translate, machine translation, neural machine translation, statistical machine translation

Research Journal in Advanced Humanities

Volume 6, Issue 4, 2025

ISSN: 2708-5945 (Print)

ISSN: 2708-5953 (Online)

ARTICLE HISTORY

Submitted: **25 November 2025**

Accepted: **06 December 2025**

Published: 07 **December 2025**

HOW TO CITE

Al-Abbas, L. S., Khalaf, B. K., Alali, O., & Alqaryouti, M. (2025). Tracking improvement from statistical to neural machine translation: An error-based evaluation of google translate. *Research Journal in Advanced Humanities,* 6(4). https://doi.org/10.58256/hnww2b65

Published in Nairobi, Kenya by Royallite Global, an imprint of Royallite Publishers Limited

1. Introduction

It is a truth universally acknowledged that the rise of globalization and advancements in technology have increased the demand for effective communication across different languages. There has been, in parallel, an increased need for translating content quickly and efficiently (khalaf & Zinb, 2020; Khalaf et al., 2022). This made professionals think of machine translation (MT) as a faster, cheaper, and more available alternative to human translators even though it may not fully replace them.

MT is the process through which a computer software transfers a text from one language into another (Okpor, 2014). It does not merely substitute words in one language with their equivalents in another; rather, it involves complex linguistic knowledge: morphology, syntax, semantics, and an understanding of concepts such as ambiguity (Al-Samawi, 2014). Although MT has gained popularity in recent years, it is argued that it will never replace human translators and will always depend on their intervention in the pre-editing and post-editing stages. However, Sycz-Opon and Galuskina (2017) state that editing MT output needs less time than translating a text from scratch. Consequently, "the use of online translation tools has increased in recent years, even among less widely spoken languages" (Selijan et al., 2011, p. 331).

Among the most popular and freely accessible online MT tools is Google Translate which provides instant translation of documents. It is widely used because it offers translation services for many different combinations of language pairs and is considered more accurate than many other MT systems (Aiken and Ghosh, 2009; Och, 2009). Google Translate was first released as a statistical machine translation service (SMT) which relied on predictive algorithms to generate translations; however, it this approach produced inaccurate or even nonsensical translations.

In 2016, Google transitioned to a neural machine translation method (NMT) which employs deep learning techniques and a wide range of linguistic resources to translate entire sentences at once. This shift made the translations sound more natural and capture the meaning and nuances across different languages. It addressed some of MT's longstanding drawbacks, including the poor readability of automated translations and incompatibility with certain languages such as Arabic. According to Turovsky (2018), the latest version of Google Translate is 60 times more precise than the old one. However, such improvements need human evaluators who can play a crucial role in developing MT technology. According to Bonnie et al. (2010, p. 809): "The fact is we have no real substitute for human judgments of translations. Such judgments constitute the reference notion of translation quality." This study aims at investigating the extent of improvement in the translation outputs produced by Google Translate after switching to the NMT method. It mainly attempts to answer the following question:

- To what extent has Google's NMT method been successful in addressing the problems faced in the SMT in Arabic-English translation?

2. Literature Review

This section includes a brief account of MT in addition to the SMT and NMT systems. It also explores Error Analysis as an approach to evaluate MT outputs and presents previous studies on MT in English-Arabic translation.

2.1 Machine Translation (MT)

MT refers to computer-assisted automated translation between two languages. Theoretically, it falls within the field of computational linguistics, which is concerned with the modeling of natural languages using computers. The concept of machine translation was first introduced in 1930 by George Artsuni and Trojanski, who proposed the earliest ideas for interpreting natural languages through mechanical or computational means. The French engineer, George Artsuni, suggested the Mechanical Brain, a device intended to translate languages. Although it was patented, it was never produced because technology was not advanced enough to support its operation (Henisz-Dostert et al., 1979). In 1936, Trojanski proposed the first comprehensive method for using computers to translate between natural languages, but his proposal also failed to be implemented successfully.

The first attempt to translate approximately 250 words using MT was carried out by Leon Dostert and Peter Sheridan in 1954, during the height of the Cold War between the United States and the Soviet Union. The experiment was considered successful and attracted substantial funding to further develop MT and explore its potential for translating human languages. Following the US government's first achievement, the ALPAC commission was established to assess the state of MT. In its 1966 report, ALPAC stated that "there is no predictable prospect of useful machine translation" (ALPAC, 1966, p.5). This conclusion had an extremely negative impact, as it significantly stalled MT research for years.

Consequently, most of the countries abandoned their MT projects except for Japan and France, which continued the research on MT applications to translate meteorological forecasts. MT research received a boost in the 1980s with the advancement of technology, and when the Internet came into existence in the 1990s, MT received further recognition. Still, as MT systems and early Internet services were costly to operate, MT tools were initially provided as a paid service.

Today, several MT platforms, including Google Translate and Microsoft Translator, offer their MT services free of charge to all end customers. The most popular service, Google Translate, was chosen for the current study because of its widespread use. It supports automatic translation of more than a hundred languages, including Arabic. The system is utilized by over 500 million people daily, processing an estimated 100 billion words daily (Google Translate, 2021). Understanding how Google Translate works requires first exploring MT methodologies that underpin this system.

2.2 Google Translate: Statistical Machine Translation (SMT) and Neural Machine Translation (NMT)

In 2006, Google Translate was launched as SMT (Och, 2009). The translation process used a computer system and was based on patterns of text without any reference to specific language rules (Turovsky, 2019).

In 2016, a major update of Google Translate occurred with the replacement of SMT by NMT (United Language Group, 2017) which produces more accurate translations (Turovsky, 2018). It works much like a neural network in that it emulates the function of a human brain, enabling the flow and processing of information across a number of layers in order to finally generate a translated output. NMT utilizes very deep algorithms for in-depth analysis of linguistic structures and focuses on translating whole sentences as units. Advantages entailed in this approach include speed, efficiency, and adaptability. With that in mind, NMT will most likely remain one of the important translation technologies in the future, and better performance and accuracy can be foreseen.

2.3 Machine Translation Evaluation (MTE): Error Analysis

Error Analysis (EA) is used to identify various types of errors taking place in MT output. This kind of assessment can indicate the strengths and weaknesses of MT systems. According to Yang (2010), the aim of EA is error detection and identification of the reasons behind linguistic failures. For many years, EA has been an integral part of MT assessment, since this approach gives insight into the limitations of a system and points out aspects that need improvement (Llitjós et al., 2005). Thus, the detection of MT errors, when comparisons are made between the MT output and human translation, has been considered important for showing how such a system can be further enhanced to yield more acceptable results (Vilar et al., 2007; Condon et al., 2010). In addition, this kind of evaluation provides critical feedback to both users and developers concerning system design, improvement, purchasing decisions, and practical use (Hutchins & Somers, 1992). Because of this, the current study employs EA to classify the errors, identify underlying causes, and then suggest solutions regarding the two MT systems tested to translate between Arabic and English.

Several MT error analysis taxonomies have been proposed, such as Flanagan (1994), Vilar et al. (2007), Frederking et al. (2004), and Farrús et al. (2010) but the most comprehensive is Costa et al.'s (2015) which is viewed as one of the most thorough tools for assessing MT quality, as illustrated in Figure 1.

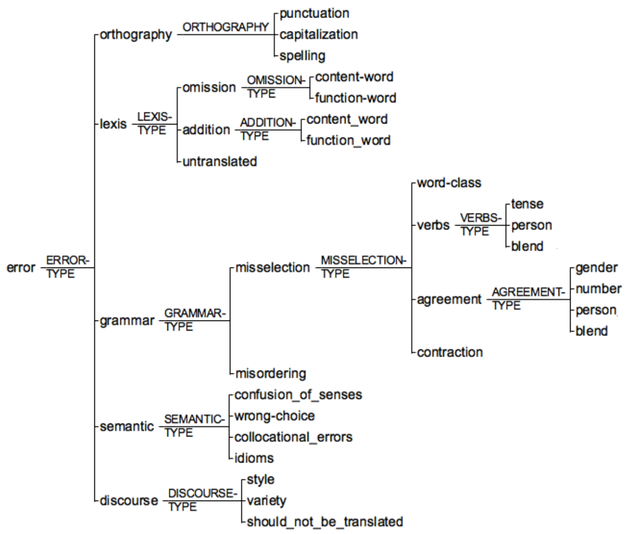

Figure 1. Taxonomy of Errors

This framework is adopted in the present study to evaluate the errors found in Google Translate after its transition to the NMT system.

2.4 Previous Studies

Since there are more people using the Internet and communication is spreading, researchers are concentrating more on Arabic works and experimenting with various methods to improve the quality of MT.

Condon et al. (2010) explored MT English-Iraqi Arabic and vice versa. Following a similar taxonomy as put forth by Vilar et al. (2006), they categorized errors under "Deletions," "Insertions," and "Substitutions" for morphological classes and kinds of errors.

Likewise, Al-Zebary (2012) investigated the challenges involved when using Google Translate, particularly those related to lexis and structure by analyzing two English and two Arabic texts which were taken from different sources. Their study found that Google translation is problematic and has many challenges concerning lexis such as (deletion, non-vocalization, addition, homographs, collocations, acronyms, and prepositions) and syntax such as (word order, subject-verb agreement and passive voice).

Similarly, ElShiekh (2012) investigated the English-Arabic and Arabic-English translation services offered by Google Translate. Three different kinds of texts were examined, namely, advertisements, literary genres, and religious texts. The study recommended not to take the final translation provided by Google Translate as it may be totally catastrophic.

Al-Samawi (2014) examined the English-Arabic Translation of 10 sentences taken from 10 encyclopaedic texts to find the errors involved in Google Translate. The study found 366 linguistic mistakes divided into three categories based on error analysis: syntactic errors (55 errors), grammar errors (174 errors), and semantic errors (137 errors).

In the same vein, Ebrahim et al. (2015) investigated the statistical methods that enhance English-Arabic machine translation. They stated that difficulties such as lexical ambiguity and pronoun resolution may arise when the target and source languages are close to each other. However, more difficulties are expected when the languages are distant such as English and Arabic and may be attributed to orthographic, morphological, and syntactical issues. They highlighted these problems with examples and came to the conclusion that English-to Arabic translation direction is highly under-presented in machine translation compared to the opposite direction. Similarly, Ali (2016) explored the problems of machine translation from Arabic into English language faced by Saudi university students of translation at the Faculty of Arts, Jazan University, Saudi Arabia. He distributed a questionnaire on the study sample which consisted of 50 randomly selected students. The study proved that students face various types of problems such as syntactic and semantic problems when using machine translation, and the rendition of target language used in the machine translator is inaccurate, especially the culture-specific terms which were out of context.

Alkhawaja et al. (2020) analyzed the translation of English texts that were rendered using NMT in Google Translate after being downloaded from broadcasting websites. The NMT outputs were assessed by three bilingual evaluators with backgrounds in linguistics and translation. The findings proved that mistranslation is the most common error type, followed by sentence meaning corruption and orthographic errors. In terms of adequacy and fluency, the translated texts are of acceptable quality.

Likewise, Almahasees et al. (2021) examined the Arabic output of the Google Translate service in English COVID-19 texts. The study found that there was a set of errors: semantic, grammatical, lexical, and punctuation. Such errors make the translated text difficult to understand. Additionally, it showed that while MT may be useful in translating basic information about COVID-19, it is still unable to handle crucial information about the virus. The study concluded that while MT can be a useful instrument, it will never fully replace human translators.

Although numerous studies have analyzed errors in MT outputs, little attention has been given to a diachronic analysis that traces how these systems have improved over the years. Therefore, this study fills this gap by examining the extent to which MT has improved and how lexical and structural errors have been reduced after shifting from SMT to NMT.

3. Methodology

In this present research, the translation problems identified by Al-Zebary (2012) when Google Translate was still operating within the SMT model were revisited. His examples were retranslated in 2023 and 2025, by which time Google Translate had already replaced SMT with NMT. The sentences were arranged in tables that included the 2012 Google Translate output (SMT), and the 2023 and 2025 Google Translate output (NMT). The translations were then compared and contrasted in order to determine the extent of improvement of the translation produced by Google Translate over the years. Such a comparative analysis used several key dimensions, including accuracy, fluency, and the system's ability to handle idiomatic expressions and context-dependent meanings.

In addition to the comparative approach, the 2023 and 2025 translations were analyzed using the error-analysis framework proposed by Costa et al. (2015), which examines translation errors across orthographic, lexical, grammatical, semantic, and discourse levels. Each NMT output was evaluated according to this model, which allowed a detailed classification of errors and a more precise identification of areas where NMT demonstrates progress or continues to struggle.

4. Findings and Discussion

English and Arabic belong to two different distant language families, namely Western Germanic and Semitic, respectively. Consequently, they are sharply different in terms of syntax, vocabulary, style, and phonology. These subtle differences impose difficulties on translators in general, and MT engines in particular. Al-Zebary (2012) investigated the problems of Google Translate in terms of lexis and structure by analyzing two English and two Arabic texts taken from different sources. His study proved that Google translation is problematic and has many challenges concerning lexis such as (deletion, non-vocalization, addition, homographs, collocations, acronyms, and prepositions) and syntax such as (word order, subject-verb agreement and passive voice). These problems are discussed in detail below, along with the 2023 and 2025 retranslations of the examples he used.

4.1 Lexical Problems

- Deletion

Al-Zebary (2012) observed that there are some cases in which Google deletes some content words from the sentences with no obvious reason. He provided examples where this problem shows clearly. However, in the 2023 and 2025 translations of the same sentences, deletion seems to have been reduced as shown in Table (1).

Table 1. Examples of deletion in Google Translate using SMT and NMT

No.	Original Text	Google Translation (SMT 2012)	Google Translation (NMT 2023)	Google Translation (NMT 2025)
1	قال الأمين العام للمنظمة الدولية بان كي مون	The Secretary General of the international organization Ban Ki-Moon	The Secretary-General of the International Organization Ban Ki-moon said	UN Secretary-General Ban Ki-moon said
2	للعام المالي الماضي	For the fiscal year	For the past fiscal year	For the past fiscal year
3	Long-term smokers die	على المدى الطويل يموت	المدخنون على المدى الطويل يموتون	يموت المدخنون على المدى الطويل

In the SMT translations, lexical deletions distorted the intended meaning and resulted in incomplete or misleading outputs. However, applying Costa et al.'s (2015) model reveals that the NMT system demonstrates substantial improvement in lexical accuracy, as the absence of omissions is a key indicator of high-quality translation. The words قال, الماضي, and smokers, which were eliminated in the SMT output, are properly rendered in the 2023 and 2025 NMT translations, which demonstrates that the new systems have greatly improved in their capability to maintain core meaning and lexical cohesion.

When the translations are considered at the discourse level, example (3) is a case in point, where NMT 2023 generated a nominal sentence: (المدخنون على المدى الطويل يموتون) which was grammatically correct but not the preferred structure in Arabic for verbal actions. In line with Costa et al.'s (2015) discourse category, the translation output should be in conformance with target language conventions. The 2025 version, however, improves this by moving to a VSO structure: (يموت المدخنون على المدى الطويل), which fits more closely the stylistic norms of Arabic. This indicates continued progress toward more natural Arabic sentence structures in the NMT system.

- Addition

Al-Zebary (2012) found that Google Translate sometimes unnecessarily added words in the translated text. The unnecessary additions disrupted the meaning, created grammatical inconsistencies, and reduced the overall accuracy of the translation. Table (2) compares the SMT output from 2012 with the NMT output from 2023 and 2025, showing how addition errors have been minimized in the newer systems.

Table 2. Examples of addition in Google Translate using SMT and NMT

No.	Original Text	Google Translation (SMT 2012)	Google Translation (NMT 2023)	Google Translation (NMT 2025)
4	واعترف بان كي مون	He admitted that Ban Ki-Moon	Ban Ki-moon acknowledged	Ban Ki-moon admitted
5	ووصف المفاوض الأمريكي جوزيف تورسيلا	He described the U.S. negotiator Joseph Trusala	He described American negotiator Joseph Torsella	American negotiator Joseph Torsella described
6	More than half of smokers underestimate	أكثر من نصف المدخنين عن التدخين يقلل	أكثر من نصف المدخنين يقللون من شأنهم	أكثر من نصف المدخنين يقللون من تقديرهم

The SMT output in example (4) adds the pronoun "He" without any linguistic reason, transforming the Arabic phrase واعترف بان كي مون into a full English clause with an overt subject. This lexical addition is unjustified and changed the sentence structure. On the other hand, NMT (2023 and 2025) significantly reduces lexical additions and renders the text more concise. According to Costa et al. (2015), avoiding unjustified lexical insertions ensures lexical fidelity, and the NMT system is clearly better in this respect. Moreover, sensitivity to the pragmatic force of words became more refined in NMT 2025. This is represented in the lexical choices of the English equivalents to اعترف, acknowledged (NMT 2023) and admitted (NMT 2025), where the former is more neutral and does not imply wrongdoing, while the latter carries a stronger and more evaluative meaning, and aligns better with the original Arabic verb. According to Costa et al. (2015), acknowledged would be a semantic weakening error, since it does not keep the required weight and tone of meaning in the original Arabic text.

Costa et al. (2015) also highlight the relevance of orthographic accuracy for MT evaluation. Example (5) is clearly an improvement across versions: the SMT output misspells the proper noun as Trusala while both NMT outputs use the correct spelling Torsella, reflecting improved named-entity recognition and greater adherence to standard orthography. Regarding the grammatical level, the translation remained problematic in NMT (2023) as Google treats the verb وصف 'describe' as having an implicit third-person masculine subject, and المفاوض الأمريكي جوزيف تورسيلا 'American negotiator Joseph Torsella' functions as its object. However, this issue appears to be reduced in the 2025 output, indicating improved handling of Arabic verbal structures and subject identification.

Example (6) illustrates semantic distortion in the SMT output due to unjustified words (عن التدخين)

that make the sentence incomprehensible. NMT-2023 and NMT-2025 remove the additions and maintain the meaning. Costa et al.'s (2015) semantic category stresses the accurate transmission of meaning, and here NMT clearly outperforms SMT by avoiding unnecessary additions and restoring semantic clarity.

- Non-vocalization

Non-vocalization arises when Arabic words are not marked with diacritics which may lead to ambiguity in Arabic-English Translation. MT engines may not trace the types of diacritics correctly, and therefore, result in mistranslation. The example that Al-Zebary (2012) gave is shown in Table (3) along with the NMT 2023 and 2025 translations.

Table 3. Example of non-vocalization in Google Translate using SMT and NMT

No.	Original Text	Google Translation (SMT 2012)	Google Translation (NMT 2023)	Google Translation (NMT 2025)
7	حتى أقدم على إزهاق روح زوجته	Even the oldest on the spirit of the loss of his wife	Until he took his wife's life	He even went so far as to kill his wife

In example (7), the SMT output from 2012 mistranslates the verb أقدم as "oldest", since the system reads it as the comparative adjective أقدمُ ("older/more ancient") rather than the verb أقدَم ("to commit, to embark on"). This kind of ambiguity reflects a core limitation of SMT: it cannot infer the correct meaning without diacritics. NMT (2023 and 2025) correctly renders the expression as "took his wife's life" or "kill his wife," showing improved lexical disambiguation. However, under the discourse category of Costa et al. (2015), NMT 2025 shows better handling of idiomatic and pragmatic force, yielding a translation closer to the norms of the target language. The construction "went so far as to" is closer to the way that English reports extreme or criminal acts, so it is more natural within English journalistic discourse.

- Homographs

Homographs are words that share the same written form but carry different meanings or grammatical functions. They usually pose a challenge because correct interpretation requires context-sensitive processing. Al-Zebary (2012) stated that homographs are problematic as the SMT system of 2012 could not resolve ambiguities, resulting in literal and semantically inappropriate outputs. The examples he gave included the word mark in English and نالت in Arabic and are listed in Table (4).

Table 4. Examples of homographs in Google Translate using SMT and NMT

No.	Original Text	Google Translation (SMT 2012)	Google Translation (NMT 2023)	Google Translation (NMT 2025)
8	Libya marks 1st independence day in 42 years	ليبيا علامات 1يوم الاستقلال في 42 سنوات	تحتفل ليبيا بعيد الاستقلال الأول منذ 42 عامًا	ليبيا تحتفل بعيد الاستقلال الأول منذ 42 عامًا
9	نالت اللعنة المجتمع الأمريكي	But won the curse of American society	Damn American society	The curse has befallen American society

SMT fails to identify the correct sense of homographs. In Example (8), the English verb "marks" is incorrectly rendered as the plural noun علامات, where SMT fails to make the distinction between the noun mark (grade, sign, label) and the verb mark meaning commemoration or celebration. Costa et al. (2015) describe this mistake as a lexical choice error where the incorrect sense of a polysemous word has been selected. The 2023 and 2025 NMT outputs correctly recognize "marks" as a verb meaning celebrates, producing the correct and natural rendering تحتفل ليبيا بعيد الاستقلال. This reflects NMT's far superior lexical disambiguation. Moreover, the 2025 NMT output points to one major enhancement in regard to adherence to the Arabic news headline convention. News headlines in Arabic typically begin with a noun, mostly the main actor or topic, rather than a verb. Such headline construction expresses a common stylistic pattern in Arabic journalism which makes prominent the performer of the action. The 2025 translation “ايبيي تحتفل بعيد الاستقلال الأول منذ 42 عامًا” follows this convention by beginning with the noun ايبيي, while the 2023 output “تحتفل ايبيي…” begins with a verb, and though grammatically sound, is stylistically less aligned with Arabic headline norms. Following Costa et al.’s (2015) discourse-level criteria, a high-quality translation should reflect not only semantic accuracy but also genre-specific stylistic features. This is indicative of a significant gain in the capacity of NMT 2025 to capture the pragmatic and stylistic expectations of the target language beyond lexical and grammatical accuracy.

Similarly, in Example (9), the Arabic verb نالت is polysemous. Its literal meaning is *obtained* or *won*, although its idiomatic meaning, which is intended here, is best rendered as *befell*, or *affected negatively*. SMT literally renders it as *won*, resulting in “But won the curse…”, which is semantically illogical. NMT 2023 uses “Damn American society”, which preserves the evaluative force, but changes the syntax. NMT 2025 improves further, producing “The curse has befallen American society”, which is fully congruent with the intended meaning and constitutes a more neutral, textually coherent rendering of the quote, that aligns with the English journalistic discourse. This reflects better lexical-sense resolution and conforms to Costa et al.’s (2015) lexical accuracy criteria.

- Collocations

Translating collocations is one of the main challenges in translation, as their meaning cannot be obtained by translating each word separately; collocations are fixed or semi-fixed multiword units. Al-Zebary (2012) explained that both human learners and MT engines make mistakes in collocations due to the difficulty of determining appropriate equivalents, as they demand contextual and idiomatic knowledge. Table (5) illustrates a comparison of how Google Translate handled collocations in SMT (2012) and NMTs (2023 and 2025).

Table 5. Examples of collocations in Google Translate using SMT and NMT

No.	Original Text	Google Translation (SMT 2012)	Google Translation (NMT 2023)	Google Translation (NMT 2025)
10	long-term smokers	طويلة الأجل المدخنين	المدخنون على المدى الطويل	المدخنون على المدى الطويل
11	long-term smokers die	على المدى الطويل يموت	المدخنون على المدى الطويل يموتون	يموت المدخنون على المدى الطويل
12	بما يوفر أموال دافعي الضرائب الأمريكيين بالملايين	Providing money U.S. taxpayers millions	Which saves American taxpayers money by the millions	This will save American taxpayers millions of dollars

SMT treats collocations as separate words, yielding literal and unnatural translations. In examples (10) and (11), “long-term smokers” becomes “طويلة الأجل المدخنين”, which is grammatically incorrect and lexically inaccurate. In contrast, both NMT outputs correctly render the phrase as “المدخنون على المدى الطويل”, which is the standard Arabic equivalent. This illustrates that NMT identifies “long-term smokers” as a collocation rather than two separate words—which is a major enhancement in lexical-sense grouping consistent with the model of Costa et al. (2015). NMT 2023 restores grammaticality with “المدخنون على المدى الطويل يموتون”, and the 2025 version further enhances word order to “يموت المدخنون على المدى الطويل”, which is closer to Arabic’s preference for verbal sentence structure. Likewise, in example (12), SMT fails to deliver the intended meaning of collocations. The expression “Providing money U.S. taxpayers millions” is semantically broken and incoherent because SMT could not reconstruct the idiomatic meaning of “save taxpayers millions”. NMT (2023 and 2025), on the other hand, correctly renders the meaning, with the 2025 version showing better structure and emphasis, closer to natural English use. The difference between 2023 and 2025 is especially salient in how collocations have been fitted into full sentences. The translation in 2023 is accurate but slightly literal or formulaic whereas the 2025 translation shows greater fluency, better word order, and clearer coherence at sentence level. This supports the fact that advanced NMT models are increasingly capable not only of handling collocations accurately, but also naturally.

- Acronyms

Acronyms are a well-recognized challenge in translation, as they demand that the system recognize abbreviated forms and retrieve their correct expanded meanings in context. Al-Zebary (2012) noted that the SMT systems usually fail to interpret acronyms and then leave them untranslated as demonstrated in Table (6).

Table 6. Examples of acronyms in Google Translate using SMT and NMT

No.	Original Text	Google Translation (SMT 2012)	Google Translation (NMT 2023)	Google Translation (NMT 2025)
13	NTC	NTC	NTC	المجلس الوطني الانتقالي
14	NHS	NHS	NHS	هيئة الخدمات الصحية الوطنية

As Table (6) shows, SMT and NMT (2023) both failed to provide any semantic content for the acronyms, leaving the translation semantically incomplete. The outputs just repeat the source acronyms, without any attempt at interpreting their meaning (NTC, NHS), leaving the Arabic reader left without any useful information. In contrast, NMT (2025) successfully expands both acronyms: المجلس الوطني الانتقالي and هيئة الخدمات الصحية الوطنية which means the 2025 model maps acronyms to their proper lexical equivalents and thus is considered a significant improvement in lexical disambiguation. It indicates that NMT models profit from continuously updated datasets to enhance their interpretive capability for abbreviations commonly used in global media and government discourses.

- Prepositions

Prepositions are one of the areas of difficulty in human and machine translation, since their meanings tend to shift with context, and direct equivalents do not always exist across languages. Al-Zebary (2012) noted that SMT tends to translate prepositions literally, which often results in semantically inaccurate or unnatural output as demonstrated in Table (7).

Table 7. Examples of prepositions in Google Translate using SMT and NM

No.	Original Text	Google Translation (SMT 2012)	Google Translation (NMT 2023)	Google Translation (NMT 2025)
15	وافقت الدول الأعضاء بالأمم المتحدة	Member states approved the United Nations	UN member states agreed	United Nations member states agreed
16	Under King Idris	تحت الملك إدريس	في عهد الملك إدريس	في عهد الملك إدريس
17	By 70,000	بواسطة 70000	بنسبة 70000	بـ 70,000

In Example (15), SMT literally translates "بالأمم المتحدة" as "Member states approved the United Nations", which is incorrect from a lexical standpoint and reverses the intended semantic relationship. The NMT outputs of 2023 and 2025 produce the proper rendering of the expression, "UN member states agreed", which conveys the intended meaning. According to Costa et al. (2015), accuracy on the lexical level requires choosing the appropriate usage-based prepositional equivalent, not form-based quality, where the NMT versions excel compared to SMT.

Similarly, in Example (16) the SMT output does not recognize that "under" here refers not to physical position but to *rule* or *authority*, leading to lexical-level failure. NMT 2023 and 2025, on the other hand, translate the phrase idiomatically as "في عهد الملك إدريس", which is the standard Arabic collocation that refers to *historical* or *political eras*. The prepositions are used in their contextual sense, hence providing better semantic accuracy.

In example (17), the SMT rendered "By 70,000" as بواسطة 70000, a grammatical structure typically used for agency in Arabic as in "written by". This ignores Arabic morphosyntactic requirements. NMT 2023 only partially corrected this, rendering it as بنسبة 70000, while the 2025 version further enhances the structure to بـ 70,000, which is more concise and natural in Arabic numerical expressions. This reflects a higher degree of translation maturity, where the system is able to go beyond the translation of mere words to recognizing the communicative function of the phrase within its genre.

4.2 Structural Problems

- Word Order

There are major differences in the word order of English and Arabic. English has verbal sentences only, whereas Arabic has both nominal and verbal sentences. Even the verbal sentences in both languages have the same components but different structures. The basic structure of the English sentence is SVO whereas the structure of the Arabic sentences is VSO. Moreover, adjectives follow nouns in Arabic but precede nouns in English.

Al-Zebary (2012) provided examples that show how these differences cause problems when translating texts from English into Arabic and vice versa as shown in Table (8).

Table 8. Examples of word order in Google Translate using SMT and NMT

No.	Original Text	Google Translation (SMT 2011)	Google Translation (NMT 2023)	Google Translation (NMT 2025)
18	بينما طالبت البلدان النامية	While demanding that developing countries	While developing countries demanded	While developing countries demanded
19	لم يكن احتلال العراق كارثياً فقط على العراق وأهله	Was not only disastrous occupation of Iraq on Iraq and its people	The occupation of Iraq was not only disastrous for Iraq and its people	The occupation of Iraq was not only disastrous for Iraq and its people

Example (18) illustrates how SMT renders "بينما طالبت البلدان النامية" as a fragment which misplaced the subject and changed the clause into a non-finite structure ("demanding") instead of a finite verb ("demanded"). NMT 2023 and 2025 correctly restore the subject–verb relationship ("While developing countries demanded"), demonstrating enhanced lexical and syntactic understanding.

Likewise, in Example (19), SMT distorts the semantic structure by fronting the adjective rather than the subject: "Was not only disastrous occupation of Iraq…". NMT (2023 and 2025) correct this by rendering the clear semantic structure: "The occupation of Iraq was not only disastrous for Iraq and its people. It is obvious that SMT followed Arabic word order too closely, leading to ungrammatical English. Both NMT 2023 and 2025 produce fully grammatical English sentences, featuring stronger mastery of English syntax. Such improvements, according to Costa et al. (2015), consist of enhancing discourse-level appropriateness to make sure the text sounds natural and logical to a native English audience.

- Subject-Verb Agreement

Subject-verb agreement is one of the challenging issues in both English and Arabic grammars, and Al-Zebary (2012) indicated that the SMT often made mistakes in number agreement because it lacked the linguistic depth to analyze complicated subjects. Table (9) clearly shows this point.

Table 9. Example of subject-verb agreement in Google Translate using NMT

No.	Original Text	Google Translation (SMT 2012)	Google Translation (NMT 2023)	Google Translation (NMT 2025)
20	وكانت الولايات المتحدة والدول الأوروبية التي تتعرض للأزمة قد ضغطت من أجل خفض الميزانية	The United States and European countries exposed to the crisis has pushed for budget cuts	The United States and European countries in crisis have pressed for budget cuts	The United States and European countries facing the crisis had pushed for budget cuts

In example (20), SMT fails to establish the appropriate agreement between the compound plural subject "The United States and European countries" and the verb. The incorrect singular form "has pushed" shows that

SMT cannot process coordinated noun phrases nor determine plurality. Both NMT outputs, 2023 and 2025, correct this by rendering the verb in the appropriate plural form: “have pressed”, “had pushed”. However, the 2025 output shows higher refinement as “had pushed” adds temporal nuance, which matches the Arabic past perfect تطغض دق. The 2025 version therefore corrects not only the grammatical structure but also captures the perfective aspect implied in the Arabic sentence.

- Passive Voice

Arabic and English do not exhibit parallel behaviour in the passive voice construction. The passive voice in English allows both agentive (with agent mentioned) and agentless passives, whereas Arabic only uses agentless passive forms, and rarely mentions the agent directly in a passive sentence. According to Al-Zebary (2012), these differences frequently cause translation difficulties, leading MT systems to produce Arabic sentences that are grammatically correct but stylistically unnatural. Table (10) provides examples that illustrate these issues.

Table 10. Example of passive voice in Google Translate using NMT

No.	Original Text	Google Translation (SMT 2012)	Google Translation (NMT 2023)	Google Translation (NMT 2025)
21	Libya was occupied for decades by various nations	احتلت ليبيا على مدى عقود من قبل الدول المختلفة	احتلت ليبيا لعقود من قبل دول مختلفة	لقد كانت ليبيا محتلة لعقود من الزمن من قبل دول مختلفة
22	Sharkas had been appointed to the same post by Qaddafi	كان قد عين في منصب شركس نفس القذافي	عين القذافي شركس في نفس المنصب	وقد تم تعيين شركس في نفس المنصب من قبل القذافي

In examples (21) and (22), the passive voice in English involves an explicit agent introduced by *by*, a structure that is acceptable in English but stylistically marked in Arabic, which generally favors agentless passives. SMT revealed strong literal transfer, reproducing نم لبق as a fixed equivalent of *by*, resulting in grammatically correct but unnatural Arabic. With the shift to NMT (2023), Google Translate presented partial improvement through the suppression of the agent in Example 22 (“نيع يفاذقلا سكرش.”), which introduces an active structure that fits better with Arabic standards; this, however, causes semantic distortion through a change in sentence focus. NMT 2025 output remains more fluent but also still manifests overexplicitness by the excessive deployment of نم لبق, a structure used in natural Arabic only in legal or extremely formal settings. It can be said that the development from SMT to advanced NMT presents improved grammaticality but continued difficulty in handling the cross-linguistic mismatch in passive voice.

5. Conclusions and Recommendations

This study demonstrated that there have been considerable improvements in lexical and structural accuracies brought about by the development of Google Translate from a simple SMT to the more advanced NMT systems in English-Arabic translation. The errors found in earlier SMT outputs, including deletion, addition, misinterpretation of homographs, incorrect collocations, literal handling of prepositions, and violations of word order in the target language, often led to distorted meaning and unnatural or incomplete translations. NMT systems, especially the 2025 version, are significantly enhanced in lexical disambiguation, orthographic precision, collocational awareness, and syntactic organization, hence achieving more coherent and contextually appropriate outputs.

Despite those advances, the results show that certain mistakes continue to exist because of fundamental differences between English and Arabic. Passive constructions and other areas of complexity continue to yield unnatural phrasing that reflect an underlying reliance on literal translation patterns. Although NMT has become increasingly sensitive to discourse factors, it still has some limitations related to capturing stylistic norms and pragmatic nuances across languages that are grammatically dissimilar. Such findings strengthen the argument that NMT, although considerably more advanced, has not yet achieved full human-like handling of cross-linguistic complexity. This lends support to Ali (2016), Alkhawaja et al. (2020) and Almahasees et al. (2021)

who found that although the adequacy and fluency of the translated texts are of acceptable quality, there are still some errors that need human editing.

This study recommends that MT developers include morphological and diacritic-based processing in their NMT models since it is a contribution towards creating a system that disambiguates Arabic homographs and helps resolve syntactic ambiguity. Genre-based training data can be included as well, particularly in areas such as journalism, law, and political discourse, which are useful in allowing MT systems to recognize stylistic conventions inherent in such examples as noun-initial headlines or agent suppression in passives. Future studies should, therefore, assess the performance of the NMT systems while processing longer texts, not just single, isolated sentences, in order to see whether the gains made at the sentence level carry over to broader discourse cohesion. Such a comparative investigation using different MT engines would be useful, too, in order to establish whether the developments under discussion constitute a field-wide trend or are specific to Google's models. Finally, further studies could also be conducted into human-machine hybrid translation workflows and thus appraise the ways in which professional translators can best integrate advanced NMT outputs and, at the same time, avoid persistent errors, especially with syntactically complex or culturally loaded texts.

Conflicts of Interest:
The authors declare no conflicts of interest.

References

Ali, A. (2016). Exploring the Problems of Machine Translation from Arabic into English language Faced by Saudi University Student of Translation at the Faculty of Arts, Jazan University Saudi Arabia. *IOSR Journal of Humanities and Social Science*, *21*(4), 55-66. https://www.iosrjournals.org/iosr-jhss/papers/Vol.%2021%20Issue4/Version-2/I2104025566.pdf

Alkhawaja, L., Ibrahim, H., Ghnaim, F., & Awwad, S. (2020). Neural machine translation: Fine-grained evaluation of Google translate output for English-to-Arabic translation. *International Journal of English Linguistics*, *10*(4), 43. https://www.ccsenet.org/journal/index.php/ijel/article/view/0/42605

Almahasees, Z., Meqdadi, S., & Albudairi, Y. (2021). Evaluation of google translate in rendering English COVID-19 texts into Arabic. *Journal of Language and Linguistic Studies*, *17*(4), 2065-2080. https://search.informit.org/doi/abs/10.3316/informit.228360028176817

ALPAC. (1966). *Language And Machines Computers In Translation And Linguistics*.

Aiken, M., & Ghosh, K. (2009). Automatic translation in multilingual business meetings. *Industrial Management & Data Systems*, *109*(7), 916–925. https://doi.org/10.1108/02635570910982274

Al-Samawi, A. M. (2014). Language errors in machine translation of encyclopedic Texts from English into Arabic: the case of Google Translate. *Arab World English Journal*, 182-211.

Al-Zebary, Y. T. A. (2012). *Lexical and Structural Ambiguity in Machine Translation* (Doctoral dissertation, Sudan University of Science and Technology). http://repository.sustech.edu/handle/123456789/5975

Bonnie, D., Snover, M., & Madnani, N. (2010). Machine translation evaluation (Chapter 5). Retrieved December 31, 2022, from
https://www.semanticscholar.org/paper/Part-5%3A-Machine-Translation-Evaluation-Chapter-5.1-Dorr-Snover/22b6bd1980ea40d0f095a151101ea880fae33049

Cheng, Y. (2019). Joint Training for Neural Machine Translation. Singapore: Springer Nature Singapore Ltd. https://doi.org/10.1007/978-981-32-9748-7

Condon, S. L., Parvaz, D., Aberdeen, J. S., Doran, C., Freeman, A., & Awad, M. (2010). *Evaluation of machine translation errors in English and Iraqi Arabic* (pp. 159–168). LREC, European Language Resources Association. https://doi.org/10.21236/ADA576234

Costa, Â., Ling, W., Luís, T., Correia, R., & Coheur, L. (2015b). A linguistically motivated taxonomy for Machine Translation error analysis. *Machine Translation*, *29*(2), 127–161. https://doi.org/10.1007/s10590-015-9169-0

Ebrahim, S., Hegazy, D., Mostafa, M. G., & El-Beltagy, S. R. (2015). English-arabic statistical machine translation: State of the art. In *Computational Linguistics and Intelligent Text Processing: 16th International Conference, CICLing 2015, Cairo, Egypt, April 14-20, 2015, Proceedings, Part I 16* (pp. 520-533). Springer International Publishing. https://link.springer.com/chapter/10.1007/978-3-319-18111-0_39

ElShiekh, A. A. (2012). Google Translate service: transfer of meaning, distortion or simply a new creation? An investigation into the translation process & problems at Google. *English Language and Literature Studies*, 2(1) 56-68. http://dx.doi.org/10.5539/ells.v2n1p56

Farrús, M., Farrús, F., Costa-Jussà, M. R., Mariño, J. B., Mariño, M., & Fonollosa, J. A. R. (2010). *Linguistic-based Evaluation Criteria to identify Statistical Machine Translation Errors*. http://translate.google.com

Flanagan, M. A. (1994). Error Classification for MT Evaluation. *Amta*, 65–72.

Frederking, R. E., Taylor, K. B., Elliott, D., Hartley, A., & Atwell, E. (2004). A Fluency Error Categorization Scheme to Guide Automated Machine Translation Evaluation. In *LNAI* (Vol. 3265). http://eprints.whiterose.ac.uk/82298/

Henisz-Dostert, B., Macdonald, R., & Zarechnak, M. (1979). *Machine Translation - Bozena Henisz-Dostert, R. R. Macdonald, Michael Zarechnak - Google Books*. https://books.google.mu/books?id=St4iXxXoIIAC&printsec=frontcover#v=onepage&q&f=false.

Hutchins, W., & Somers, H. (1992). *An introduction to machine translation*.

Khalaf, B. K., & Zinb, Z. M. (2020). Developing a Transcultural Tool for Teaching Cultures in Translation Studies. *International Journal of Innovation, Creativity and Change*, *11*(8), 573-593.

Khalaf, B. K., Zin, Z. M., & Al-Abbas, L. S. (2022). Contemporary perspective on cognitive development: Reconceptualising situational context as embedded model. *International Journal of Instruction*, *15*(1), 401-420.

Llitjós, A. F., Carbonell, J. G., & Lavie, A. (2005b). *A Framework for Interactive and Automatic Refinement of Transfer-based Machine Translation* (Vol. 87).

Google Translate. (2021). *Google - About Google, Our Culture & Company News*. https://about.google/?hl=en

Och, F. (2009). *51 Languages in Google Translate*. Google Research Blog on August 31. Retrieved from https://ai.googleblog.com/2009/08/51-languages-in-google-translate.html

Okpor, M. D. (2014). Machine translation approaches: issues and challenges. *International Journal of Computer Science Issues (IJCSI)*, *11*(5), 159.

Sycz-Opoń, J., & Gałuskina, K. (2017). Machine translation in the hands of trainee translators–An empirical study. *Studies in Logic, Grammar and Rhetoric*, *49*(1), 195-212. https://sciendo.com/article/10.1515/slgr-2017-0012

Selijan, S., Brkic, M. & Kucis, V. (2011). *Evaluation of free online machine translations for Croatian-English and English-Croatian language pairs*. Paper presented at 3rd International Conference of The future of Information Science (INFuture) in Zagreb, 9-11 November 2011. pp. 331-340.

Turovsky, B. (2017). *Higher quality neural translations for a bunch more languages*. The Keyword Google Blog. Google. Retrieved March 6, 2019, from https://www.blog.google/products/translate/higher-quality-neural-translations-bunch-more-languages/

Turovsky, B. (2018). *Email interview: D'Monte, Leslie: Working to improve quality of translations*. Google's Barak Turovsky in EmTech India 2018—an emerging technology conference organized by Mint and MIT Technology Review. Retrieved January, 2023, from https://www.livemint.com/Companies/hrQAcRhDJm6acbKrwLKtlN/Working-to-improve-quality-of-translations-says-Googles-Ba.html

Vilar, D., Xu, J., Fernando D'haro, L., & Ney, H. (2007). *Error Analysis of Statistical Machine Translation Output*. http://www.tc-star.org

Yang, W. (2010). A Tentative Analysis of Errors in Language Learning and Use. *Journal of Language Teaching and Research*, *1*(3), 266–268. https://doi.org/10.4304/jltr.1.3.266-268