OPEN ACCESS
Received: 20 -10 -2024
Accepted: 01- 01-2025

الآداب
للدراسات اللغوية والأدبية

Machine Translation Evaluation between Arabic and English during 2020 to 2024: A Review Study

Aziz Mohammed Abdo Saeed *
aziz.saeed@hotmail.com

Abstract:

The field of machine translation is one of the new fields of study that attracts the interest of many researchers. The evaluation of machine translation is central to the field and so many research has been conducted on this. This study presents a review of the studies conducted on evaluation of machine translation between Arabic and English languages during the period 2020 to 2024. It reviews a collection of 11 studies conducted during 2020 to 2024. It presents an overview of these studies indicating their objectives, methodologies, findings. By synthesizing these studies, reader can have clear overview of current research on this area. This study presents a database for future researchers interested in machine translation, particularly evaluation of machine translation quality. The study found that Google Translate was the subject of evaluation for most of the studies and that all studies almost used human metrics for evaluation. In addition to presenting the current trends of research and a summary of a group of studies, this study suggests areas which future research should address to fill the gaps that have been noticed in the review.

Keyword: Machine Translation, Machine Translation Evaluation, Google Translate, Artificial Intelligence, Machine Translation between Arabic and English.

* Master of English Linguistics, English Department, Faculty of Arts, Thamar University, Republic of Yemen.

Cite this article as: Saeed, A. M. A. (2025). An Analysis of Syntactic Complexity in Texts Translated by Google Translate and Human Translator from Arabic into English, *Arts for Linguistic & Literary Studies*, *7*(2): 665 -678. https://doi.org/10.53286/arts.v7i2.2528

الآداب
للدراسات اللغوية والأدبية

OPEN ACCESS
تاريخ الاستلام: 2024/10/20م
تاريخ القبول: 2025/01/01م

تقييم الترجمة الآلية بين اللغة العربية واللغة الإنجليزية في المدة من 2020م إلى 2024م: دراسة مراجعة

عزيز محمد عبده سعيد *
aziz.saeed@hotmail.com

ملخص:

تعد الترجمة الآلية إحدى المجالات الحديثة التي حظيت باهتمام كثير من قِبل الباحثين، ويمثل تقييم جودة الترجمة الآلية موضوعًا جوهريًا في مجال الترجمة الآلية؛ ولذلك فقد تمت دراسات كثيرة حديثا حول هذا الموضوع. وتهدف هذه الدراسة إلى تقديم مراجعة للدراسات التي كان الهدف منها تقييم جودة الترجمة الآلية بين اللغتين العربية والإنجليزية خلال المدة من العام 2020 إلى 2024م. وتراجع الدراسة إحدى عشرة دراسة، وتقدم هذه الدراسة ملخصًا عن مجموعة من الدراسات، بالإشارة إلى أهدافها، ومنهجية التقييم فيها، والنتائج لهذه الدراسات. ومن خلال جمع هذه الدراسات يستطيع القارئ الحصول على موجز حول الدراسات الحديثة التي تمت حول هذا الموضوع. وفضلًا عن ذلك تقدم الدراسة قاعدة بيانات للباحثين المهتمين بعمل أبحاث مستقبلية عن الترجمة الآلية لاسيما تقييم الترجمة الآلية، ووجدت الدراسة أن ترجمة جوجل حظيت بالاهتمام الأكبر، وأن جميع الدراسات تقريبًا استخدمت معايير تقييم بشرية. فضلًا عن تقديم الاتجاهات والميول الحالية للباحثين، فإن هذه الدراسة تقترح موضوعات للبحث المستقبلي لملء الفجوات التي تمت ملاحظتها خلال المراجعة، وسوف يكون لذلك أثر في تحسين أداء الترجمة الآلية بين اللغتين العربية والإنجليزية.

الكلمات المفتاحية: الترجمة الآلية، تقييم الترجمة الآلية، الذكاء الاصطناعي، الترجمة الآلية بين اللغة العربية والإنجليزية.

* ماجستير لغويات، قسم اللغة الإنجليزية، كلية الآداب، جامعة ذمار، الجمهورية اليمنية.

للاقتباس: سعيد، ع. م. ع. (2025). تقييم الترجمة الالية بين اللغة العربية واللغة الانجليزية في المدة من 2020م إلى 2024م: دراسة مراجعة، *الآداب للدراسات اللغوية والأدبية*، *7*(2): 665-678. https://doi.org/10.53286/arts.v7i2.2528

1- Introduction

Machine Translation (MT) has established itself as an important technology in crossing language barriers, particularly between languages from different family origin as Arabic and English. The fast-growing globalization and the need for effective cultural communication have encouraged experts to develop and improve MT systems that make the translation of texts and speech from one language to another easier. Arabic which is a member of the Semitic language family, presents many challenges for MT to translate into English due to the complex morphology, syntax, and semantics, which are different from English which belongs to an Indo-European languages' family (Zakraoui et al., 2020). Moreover, the translation of cultural terms causes difficulties even for human translators (Mounassar, 2018). In addition to cultural challenges, the linguistic differences between English and Arabic create translation challenge (Alnasery, 2024).

Machine Translation has witnessed development of many methodologies based on different approaches, ranging from rule-based systems to statistical and neural approaches. Each approach has its points of strength and weakness. As the need for Machine Translation systems increase, it becomes increasingly important to improve evaluation methods. The good evaluation helps the developers to make more effective systems and also helps human translators as it makes their jobs easier (Ali, 2020, Elmahdi, 2024). However, the role of human translator is crucial, particularly in multicultural communication (Ashuja'a & Jibreel, 2024). Translation profession helps fostering cultural exchange and cooperation (Elmahdi & Mohamad, 2024, Motair, et al, 2025).

This paper aims to provide a comprehensive review of the studies conducted during the period from 2020 till 2024 about evaluation of machine translation systems between Arabic and English. It will explore the objectives, findings and methodologies used in each study.

Furthermore, the paper will discuss the challenges specific to Arabic-English translation, including issues related to dialectal variations, cultural context, and the intricacies of Arabic grammar as concluded in these studies. By synthesizing existing studies and their evaluation practices, this survey aims to contribute to the ongoing discourse on guiding researchers to choose new areas for future studies on machine translation evaluation and to check the studies from one source. It will help supervisors of research to check originality of the studies they supervise. This study will search for gaps in research and suggest them for future studies.

The study seeks to answer the following questions:

1- What are the points of redundancy in the studies conducted in the period 2020 to 2024 to machine translation evaluation between the pair of languages Arabic and English?
2- What were the objectives, methodology and findings of each of these studies?

3- What are the gaps that can be filled by new research on evaluation of machine translation between Arabic and English?

There has been a redundancy and repetition in some aspects in machine translation evaluation studies. For this, it was necessary to make a review of a group of studies to identify the points of redundancy in order to be avoided in future studies. In summary, this review will not only explore the current state of MT evaluation studies between Arabic and English but also identify gaps in research and suggest future directions for studies aiming at enhancing the quality of machine translations in this linguistically rich and diverse context.

2- Literature Review

The field of Machine Translation (MT) has developed largely over the last decade, particularly in relation to Arabic and English language pairs and other pairs of languages which don't belong to the same family origin. This literature review presents key studies and findings related to the evaluation of MT systems, focusing on Arabic-English evaluation studies. It discusses methodologies, challenges, and advancements in the evaluation of Arabic-English machine translations.

Evaluation Methodologies

It is crucial to assess the performance of MT systems in order to guide developers. Evaluation is a key issue for developing and improving these systems. There are various methodologies that have been suggested and developed by scholars, including the human-metrics evaluations (performed by human) that use human knowledge and cultural background and the computerized-metrics evaluation methods which are performed automatically. According to Han (2016), the evaluation performed by human usually assesses intelligibility, fidelity, fluency, and adequacy. There are also advanced methods of human evaluation including task-oriented measures and post-editing assessments.

The automatic evaluation methods, such as BLEU, METEOR, and AL-BLEU, have achieved popularity for the efficiency and scalability they have. BLEU, for example, compares between machine-generated translations and reference translations by humans for the same texts, providing a quantitative score that reflects translation quality (Zakraoui et al., 2020). However, the automatic metrics has been criticized as not reliable because of their limited ability to capture and convey the semantic, cultural and contextual nuances, leading to a call for more comprehensive evaluation methods (Han, 2016).

Challenges in Arabic-English MT

The complex nature of Arabic language creates many challenges for systems of MT. The rich morphology, varieties of dialects, and the complicated syntactic structures make the translation process difficult. Zughoul and Abu-Alshaar (2005) concluded that the variation of dialects and the morphological

richness of Arabic language make it necessary to adopt specialized approaches to MT. Sawaf (2010) emphasized that dialectal variation of language should be standardized to make the translation by machine more accurate.

Moreover, there are no parallel corpora for the rich Arabic dialects and that makes the development of effective MT systems harder. As Harrat et al. (2019) pointed out, data-driven approaches encounter challenges because of the insufficient resources, which makes training of statistical models limited. Moreover, the evolution of dialects and the appearance of Arabizi (a blend of Arabic and Latin script) creates additional challenges for MT systems (Harrat et al., 2019).

Recent development in MT Evaluation

Recent advancements in the evaluation of machine translation have focused on integrating deep learning technique into evaluation metrics. Han (2016) suggested that Automatic evaluation methods may be categorized into lexical similarity and linguistic features methods. He added that incorporating syntactic and semantic analysis can improve evaluation accuracy. The launch of neural machine translation (NMT) has also changed machine translation quality as many studies indicated that NMT systems often outperform the other approaches which are based on statistical methods (Zakraoui et al., 2020).

Furthermore, the advancements of hybrid approaches of machine translation that uses rule-based and statistical methods together has shown success in solving many of the challenges in machine translation of Arabic language into other languages.

3- Methodology

This paper uses a comprehensive survey method to review the various studies on machine translation evaluation between Arabic and English languages in the years 2020 to 2024. The goal of the study is to provide summaries of these studies and to provide synthesis of the objectives, methodologies, findings presented in the existing literature on machine translation evaluation during the five years 2020 to 2024.

Selection and Inclusion Criteria

A systematic search for the existing literature was done to identify the relevant studies on machine translation evaluation. The search used various databases such as Google Scholar, Semanticscholar, IEEE Xplore, SpringerLink, etc. Inclusion criteria focused on studies published between 2020 and 2024, ensuring that the studies aimed at evaluating machine translation and the languages pair used in the studies are Arabic and English.

Studies were included if they addressed evaluating quality of machine translation output in any of the quality aspects; accuracy, fluency, adequacy...etc. Studies that did not address evaluation of machine translation directly or were conducted before 2020 were excluded from the survey.

Data Extraction

The information extracted from each study includes the objectives, evaluation methodology, machine translation tool, the corpus used as text samples for evaluation, the direction of translation (either from Arabic into English or from English into Arabic or both directions) and the findings. This data was collected by reviewing the studies in order to provide a database that can be used for educational and further research implications.

This methodology ensures a comprehensive review of the current landscape in machine translation evaluation between Arabic and English. It offers valuable insights for researchers and practitioners of machine translation.

4- Results

This survey on studies about Machine translation evaluation between Arabic and English during the period 2020 to 2024 indicate that this area has attracted many research and many studies were conducted about it. In this section, 11 studies are reviewed and their objectives, methodologies and findings are summarized as below.

Studies Surveyed

Abdelaal and Alazzawie (2020) conducted a study aiming at identifying the types of errors in Google Translate output in translating informative news from Arabic language into English language in order to assess the fluency and the semantic adequacy and the extent to which a human translator in needed for correcting the errors. The data was analyzed using a mixed approach as the errors were quantitatively and qualitatively analyzed. The qualitative method used in the study was that proposed by Hsu's (2014) machine translation errors Classification. Localization Quality Evaluation and Multidimensional Quality Metrics were used to make the quantitative description of errors. A questionnaire was used to assess the semantic adequacy. The analysis showed that omission is the most common semantic error occurred in the output. It added that the error of inappropriate lexical choice semantic error is sometimes resulted from homophonic nature of source text words. The study concluded that machine translation can be used to make translation process faster, but it requires human intervention to make the translation accurate.

Trying to evaluate and compare three machine translation services; Google Translate, Microsoft Bing, and Ginger in translating from English into Arabic, Ali (2020) used texts from UN records as corpus to evaluate fidelity and intelligibility of the outputs of the three machine translation services using two scale scores "correct or incorrect". The quantitative evaluation marked the number of fidelity and intelligibility errors and their percentages. The study revealed that none of these services provided a perfect translation of the source texts. However, the study concluded that Microsoft Bing was rated the best while Google Translate was rated

the least accurate based on the number of fidelity and intelligibility errors. Ginger was rated between Microsoft Bing and Google Translate. The study implied that these services may be used to get gist translation of English texts into Arabic, but the accurate translation with full understanding requires post-editing of the output.

Almahasees et al., (2021a) conducted a study aiming to evaluate the adequacy and Fluency of Facebook Translation Service (FTS) in translating English posts into Arabic. The study used Translation Automation User Society (TAUS) scales which used native speakers of Arabic and near-native speakers of English as evaluators to rate the data from 1 to 4. The study revealed that FTS provided fluent output rated good which is equal to 3 on the scale, where "the output is assessed as flowing smoothly with minor linguistic errors". Moreover, the study revealed that FTS succeeded in providing an adequate output rated as "most" which is equal to 3 on a 1-4 scaling score. It revealed that the full meaning of the source text was deemed to be transferred to the target language .

Almahasees et al., (2021b) tried to scrutinize the capacity of Google Translate in rendering English texts about COVID-19 into Arabic. The study adopted human methodology in analyzing the output of Google Translate. The text samples were extracted from different health organizations corpus. The study revealed that Google Translate made different errors; in semantics, grammar, lexical and punctuation. The study concluded that MT may be a good Translate tool but cannot replace human translators.

In a similar vein, Kadaoui, et al., (2023) assessed the proficiency of "instruction-finetuned" machine translation systems; ChatGPT and Bard, in translating ten Arabic varieties. The varieties covered Modern Standard Arabic, Classical Arabic, and other several dialectal Arabic varieties. The analysis indicated that these instruction-finetuned machine translation systems encountered challenges in translating dialects, but they perform better than other "existing commercial systems". However, in translating Classic Arabic and Modern Standard Arabic they trailed behind commercial systems like Google Translate. The study also scrutinized the efficacy of Bard translation system in following human translation instructions and indicated that Bard had limited capability in recognizing context.

Khoshafah (2023) evaluated the accuracy of ChatGPT translation from Arabic into English and from English into Arabic. Texts samples were selected from different genres and the evaluation was based on comparison with professional translation for the same texts translated by ChatGPT. The methodology adopted was a comparison of semantic content of ChatGPT output with professional translations of the same texts. There was no specific methodology proposed by some scholar or researcher. The research concluded that ChatGPT can be used for simple content translation, but it struggles with complex content and thus needs human intervention.

Nagi (2023) evaluated Google Translate as an example of neural machine translation (NMT). The study used about 180 sentences from Arabic and English containing relative clauses to evaluate both directions of translation; English to Arabic and Arabic to English. Manual Error annotation was performed by professional translators in additional to BLUE automatic evaluation. The study revealed that errors related to fluency are more frequent than accuracy errors. It revealed also that quality of machine translation is lower when translating from English into Arabic than machine translation quality when translating from Arabic into English. The study concluded that the gap between machine translation quality and professional human translation quality was still large.

A study by Qassem & Aldaheri (2023) sought to assess the effectiveness of Google Translate in translating dialogue acts in English conversations into Arabic. The approach of assessment is based on Austin's theory of speech acts. The study highlighted certain challenges in identifying the performative functions of utterances in English conversations when Translated into Arabic by Google Translate. The study added that such challenges result from linguistic features of English conversations like idiomatic and polysemous words.

Aldawsari (2024) examined the advancements of AI-driven machine translation by comparing it to traditional machine translation services. It evaluated and compared Bing AI translation service to Google Translate and Bing Translator. The study used colloquial expressions from Arabic language to be translated into English using these translation systems. The study used the same Arabic sentences used in At-tall's 2019 thesis which compared Google Translate output to human translations. Using the same sentences as a test dataset allowed the study to indicate the advancements by comparing the output of 2019 to the output of 2024. The study indicated limited improvements in Google Translate " with Bing Translator exhibiting a similar level of translation accuracy" (P. 237). The study also concluded that Bing AI Chat outperformed the other two systems, "Notably, Bing AI Chat provided interpretations and valuable comments on the tested Arabic phrases, demonstrating a deeper understanding of the intended meaning" (P.237).

Another study investigated the quality of Google Translate and ChatGPT in translating scientific texts from English into Arabic. The study used error analysis based on the typology of errors introduced by Multidimensional Quality Metrics (MQM). Another evaluation is provided to the translated texts using a 7-point Likert scale. The study concluded that Google Translate outperformed ChatGPT. However, the study concluded that both systems still require a lot of training (Alzain et al., 2024).

Saeed (2024) assessed Google Translate performance in translating Arabic texts into English. The study used corpus from a famous Arabic book called the MOQADDIMAH by Ibn Khaldun and it used a reference English translation by Franz Rosenthal. The study analyzed semantic errors and used the approach of comparing semantic components of the machine translation output to the reference human translation. The

study revealed that Google Translate output contained errors of semantic concepts and semantic relations and indicated their percentages. The study also revealed that Google Translate output had higher syntactic complexity than human translation for those sample texts.

The tables 1 and 2 below show a summary of these 10 studies surveyed. Table 1 lists the machine translation systems targeted in the evaluation studies, the texts used as samples in evaluation and the direction of translation; either from Arabic to English or from English to Arabic. Table 2 lists the aspect(s) of quality evaluated in each study and either the study used human or automatic metrics in evaluation process and findings of each study.

Table 1 of Targeted Systems, Corpus Used and Direction of Translation in Each Study

Study	MT system targeted	Corpus used	Direction of Translation
(Abdelaal and Alazzawie, 2020)	Google Translate	informative news	Arabic into English
Ali, 2020)(	Google Translate, Microsoft Bing, and Ginger	Texts from UN records	English into Arabic
(Almahasees et al.,2021a)	Facebook Translation Service	Covid-19 Facebook posts	English into Arabic
(Almahasees et al.,2021b)	Google Translate	Health Organizations corpus	English into Arabic
(Kadaoui, et al., 2023).	Bard, ChatGPT & Google Translate	Modern Standard Arabic & Arabic Dialects	Arabic into English
Khoshafah, 2023)(	ChatGPT	Variety of genres	Arabic into English
Nagi, 2023)(	Google Translate	Sentences in both Arabic and English languages containing relative clauses	Arabic into English & English into Arabic
(Qassem & Aldaheri, 2023)	Google Translate	Dialogue conversations	English into Arabic
Aldawsari, 2024)(	Bing AI, Google Translate & Bing Translator	Colloquial expressions	Arabic into English
(Alzain et al., 2024).	GT and ChatGPT	Scientific texts	English into Arabic
Saeed, 2024)(	GT	Classic Arabic texts	Arabic into English

Table 2 of Aspect(s) of MT Quality, Method of Evaluation and Findings of Each Study.

Study	Aspect of MT quality evaluated	Human Evaluation or Automatic evaluation	findings
(Abdelaal and Alazzawie, 2020)	fluency and adequacy	Mixed Approach	High rate of omission error

			Inappropriate lexical choice Requirement of human intervention
Ali, 2020)(	fidelity & intelligibility	Human evaluation	None of the systems provided perfect translation Microsoft Bing is rated the best Google Translate was rated the lest accurate These systems can be used for gist translation
(Almahasees et al.,2021a)	Adequacy & fluency	Human evaluation	Output flowing smoothly with miner linguistic errors The full meaning of the source is deemed to be transferred into the target language
(Almahasees et al.,2021b)	Translation efficiency	Human	Google Translate made many errors MT may be used a good tool, but it cannot replace human
(Kadaoui, et al., 2023).	Translation proficiency	Human	Chatgpt performed better than GT in translating dialects GT performed better than Chatgpt in translating standard Arabic
Khoshafah, 2023)(	Accuracy	Human	ChatGPT may be used for simple content translation, but it struggles with complex content
Nagi, 2023)(	Fluency & accuracy	Mixed approach	Fluency errors are more frequent than accuracy errors. English to Arabic translation quality is lower than Arabic to English
(Qassem & Aldaheri, 2023)	Effectiveness of translating dialogues	Human	GT faced challenges in identifying performative functions of utterances
Aldawsari, 2024)(	Accuracy	Human	GT has improved since 2019. Bing AI outperformed GT and Bing Translator
(Alzain et al., 2024).	Accuracy	Human	Google Translate outperformed

			ChatGPT in translating scientific texts
Saeed, 2024)(	accuracy	Human	GT contained errors of semantic concepts more than semantic relations. GT showed higher syntactic complexity level

5- Discussion

The evaluation of machine translation is essential to translation studies, as demonstrated by the large number of studies conducted. However, more studies about pre-editing and post-editing are vital to guide machine translation users to get the best advantage. Machine translation between Arabic and English is complex as the pair of languages belongs to different families of languages. Throughout the studies explored, there is a clear focus on Google Translate as most of the studies targeted evaluating it either separately or in comparison to other systems. It is important at this stage to pay more attention to systems which are based on artificial intelligence technology. The corpus (text samples) used in the studies were almost varied, excluding literary texts because of the special esthetic nature. Although automatic metrics of machine translation evaluation are available and faster than human evaluation, the human evaluation is still preferable to most researchers because of its ability to account for subjective factors. Human evaluation showed deeper understanding of translation quality. The automatic metrics are used sometimes with human evaluation.

A notable finding is that machine translation systems show better performances in translating scientific texts, while they struggle in translating dialects and non-standard texts.

6- Conclusion and Implications

The recent years noticed a notable trend and growing interest in evaluating machine translation systems as indicated in the studies explored in this paper. Google Translate system attracted most of the interest showed by researchers, in addition to the AI-based systems that has showed better performance in understanding cultural nuances as showed in the studies explored.

This survey provides a vital database for researchers on machine translation evaluation as it provides a synthesis of objectives, methodologies and findings for a group of recent studies conducted during the last 4 years. This will guide researchers to choose a new topics and new areas in machine translation in general and machine translation evaluation in particular. This database will also serve as a collective source for new studies.

7- Recommendations for Further Studies

Based on exploring the group of studies targeted in the survey, it is recommended for new researchers to conduct studies on pre-editing and post-editing of machine translation output. It is important to integrate evaluation studies with pre-editing and post-editing processes in machine translation as this will provide insights to the linguistic features that make a text translatable in machine with lest possible errors. This can lead to a style of writing that can be translated by machine with more accuracy and consequently leads to ease of cross-cultural communication.

It is recommended also that researchers conduct more experimental studies that can provide practical solutions to many challenges faced by machine translation. Researchers from different families, like Arabic and English, are recommended to cooperate in studying and analyzing the linguistic features that cause such challenges. This will provide insights to developers of machine translation systems.

Reference

Abdelaal, N. M., & Alazzawie, A. (2020). Machine Translation: The Case of Arabic-English Translation of News Texts. *Theory & Practice in Language Studies (TPLS)*, 10(4). Page range

Aldawsari, H. A. (2024). Evaluating Translation Tools: Google Translate, Bing

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* (05), 524–548. https://doi.org/10.4236/ojml.2020.105030

Almahasees, Z., Al-Taher M., &, Helene J. (2021a). Evaluation of Facebook Translation Service (FTS) in Translating Facebook Posts from English into Arabic in Terms of TAUS Adequacy and Fluency during Covid-19. *Advances in Science, Technology and Engineering Systems Journal 6*(1) 1241-1248 .https://dx.doi.org/10.25046/aj0601141

Almahasees, Z., Meqdadi, S., & Albudairi, Y. (2021b). Evaluation of Google Translate in Rendering English COVID-19 Texts into Arabic. *Journal of Language and Linguistic Studies, 17*(4), 2065-2080. Doi: 10.52462/jlls.149

Alnasery, I. H. S. (2024). Reducing the Ambiguity in Translating Prepositions from English into Arabic, *Arts for Linguistic & Literary Studies, 6*(2): 541 -555

Alzain E., Nagi K.A., Algobaei, F. (2024). The Quality of Google Translate and ChatGPT English to Arabic Translation: The Case of Scientific Text Translation. *Forum for Linguistic Studies. 6*(4): 837-849. DOI: https://doi.org/10.30564/fls.v6i3.6799

Ashuja'a, A., & Jibreel, I. (2024). Translator Praxis: An Investigation into the Practical Component in BA Translation Programs at Yemeni Universities. *Arts for Linguistic & Literary Studies*, *6*(3), 574–604. https://doi.org/10.53286/arts.v6i3.2085

At-tall, S. M. (2019). Comparative Study between Google Translator and Human Translator in Rendering Colloquial Arabic Expressions in the Late Prime Minister Wasfi At-Tall's Speeches into English, (Unpublished Master's thesis). Yarmouk University, Irbid, Jordan.

Elmahdi, O. E. H., & Mohamad, H. M. H. (2024). Developing Translation and Interpretation Skills: An Analysis of Students' Linguistic Needs. *Arts for Linguistic & Literary Studies*, *6*(4), 652–671. https://doi.org/10.53286/arts.v6i4.2201

Han, L. (2016). Machine Translation Evaluation Resources and Methods: A Survey. *arXiv preprint arXiv:1605.04515*.

Harrat, S., Meftouh, K., & Smaili, K. (2019). Machine Translation for Arabic Dialects (Survey). *Information Processing & Management, 56*(2), 262-273.

Hsu, J. A. (2014). Error Classification of Machine Translation A Corpus-based Study on Chinese-English Patent Translation. *Journal of Translation Studies*, 121-136.

Kadaoui, K., Magdy, S. M., Waheed, A., Khondaker, M. T. I., El-Shangiti, A. O., Nagoudi, E. M. B., & Abdul-Mageed, M. (2023). Tarjamat: Evaluation of Bard and Chatgpt on Machine Translation of Ten Arabic Varieties. *Preceedings of the First Arabic Natural Language Processing Conference*, ArabicNLP 2023, 52-75. https://doi.org/10.18653/v1/2023.arabicnlp-1.6

Khoshafah, F. (2023). *ChatGPT for Arabic-English Translation: Evaluating the Accuracy*. ResearchSquare. https://doi.org/10.21203/rs.3.rs-2814154/v2

Motair, A. A. A., Algobaei, F., & Alhazmi, M. D. (2025). Ethics in Translation: A Pathway to Integrity in Future Professionals. *Arts for Linguistic & Literary Studies*, *7*(1), 711–732. https://doi.org/10.53286/arts.v7i1.2417

Mounassar, A. (2018). Difficulties and Problems Facing English Students in Translating Culture-Specific Items from English to Arabic and their Solutions. *Journal of Arts, 1*(7), 496–437. https://doi.org/10.35696/.v1i7.520

Nagi, K. A. (2023). Arabic and English Relative Clauses and Machine Translation Challenges. *Journal of Social Studies, 29*(3), 145–165. https://doi.org/10.20428/jss.v29i3.2180

Qassem, M. & Aldaheri, M. M. (2023). Can Machine Translate Dialogue Acts: Evidence from Translating Dialogues from English to Arabic. 3L: *Language, Linguistics, Literature,29* (4), 63-81

Saeed, A. M. (2024). Quality Assessment of Google Translate: Translating Texts from Arabic into English [Unpublished Master's thesis]. Thamar University, Thamar, Yemen.

Sawaf, H. (2010). Arabic dialect handling in hybrid machine translation. In *Proceedings of the 9th Conference of the Association for Machine Translation in the Americas: Research Papers*.

Translator, and Bing AI on Arabic Colloquialisms. *Arab World English Journal (AWEJ) Special Issue on ChatGPT*, April 2024: 237-251.DOI: https://dx.doi.org/10.24093/awej/ChatGPT.16

Zakraoui, J., Saleh, M., Al-Maadeed, S., & Alja'am, J. (2020). Evaluation of Arabic to English Machine Translation Systems. *11th International Conference on Information and Communication Systems (ICICS),* 185-190. https://doi.org/10.1109/ICICS49469.2020.239518.

Zughoul, M. R. & Abu-Alshaar, A. M. (2005). English/Arabic/English Machine Translation: A Historical Perspective. *Meta, 50*(3), 1022–1041. https://doi.org/10.7202/011612ar