Machine Translation in Scientific Texts : Balancing Accuracy and Content

Buraq MELHEM[1*], Mutasim SHUNNAQ[2], Rola SHLOOL[3]

[1] ***Jadara University, Jordan***
boraq511@gmail.com

[2] ***Jadara University, Jordan***
motasim.b@jadara.edu.jo

Received date: 05/06/2025 **Revised date: 06/06/2026** **Publication date: 01/10/2026**

Abstract :

This study analyzes the role of machine translation within the field of scientific texts, that is, with regard to issues of terminological accuracy and contextual appropriately. With computerized translation software increasingly utilized in scholarship and research environments, there are concerns regarding such software's ability to offer accurate, reasonable, and contextually appropriate translations—especially where terminology transfers conceptual relevance. Scientific texts are not merely linguistic constructs but are also embedded in theoretical and methodological contexts that require more than literal translation ; they require nuanced sensitivity to both language and discipline knowledge. The research seeks to evaluate the performance of widely used machine translation tools such as Google Translate and DeepL in scientific text translation in many fields such as biomedical sciences, physics, and artificial intelligence. It considers how these systems handle technical vocabulary, subtle syntactic organization, and situational meaning that is constitutive of the integrity of scientific discourse.

Keywords : Machine Translation ; Scientific Text ; Context ; Accuracy ; Scientific Terminology ; Google Translate.

ملخص:

تُحلل هذه الدراسة دور الترجمة الآلية في مجال النصوص العلمية، من حيث دقة المصطلحات وملاءمتها للسياق. مع تزايد استخدام برامج الترجمة المحوسبة في بيئات البحث العلمي، تُثار مخاوف بشأن قدرتها على تقديم ترجمات دقيقة ومعقولة وملائمة للسياق، لا سيما عندما تُضفي المصطلحات

[1] **Corresponding author.** Buraq MELHEM, boraq511@gmail.com

أهمية مفاهيمية. فالنصوص العلمية ليست مجرد بنى لغوية، بل هي أيضًا جزء لا يتجزأ من سياقات نظرية ومنهجية تتطلب أكثر من مجرد ترجمة حرفية؛ فهي تتطلب حساسية دقيقة لكل من اللغة والمعرفة التخصصية. يسعى البحث إلى تقييم أداء أدوات الترجمة الآلية واسعة الاستخدام، مثل Google Translate وDeepL، في ترجمة النصوص العلمية في العديد من المجالات، مثل العلوم الطبية الحيوية والفيزياء والذكاء الاصطناعي. ويتناول كيفية تعامل هذه الأنظمة مع المفردات التقنية، والتنظيم النحوي الدقيق، والمعنى الظرفي الذي يُشكل جوهر الخطاب العلمي.

الكلمات المفاتيح: الترجمة الآلية؛ النصوص العلمية؛ السياق؛ الدقة؛ المصطلحات العلمية؛ ترجمة جوجل.

****

Introduction :

During the past two decades, enormous growth has gone into machine translation, which has evolved from fairly simple direct dictionary-style translation into more sophisticated AI-based systems such as statistical and neural translation. Tools like Google Translate and DeepL have become household names in translating all kinds of texts for laymen, academics, and professionals. With this growing use, fundamental questions arise regarding the appropriateness of such tools to specialized texts, especially scientific texts that are highly terminologically accurate and context-sensitive. A scientific text is measured not only by the correctness of its linguistic structures but also by its ability to transmit into another language the precise meaning that maintains the conceptual load of the terms and their context in the cognitive framework of researchers or scientists. Words in this kind of text are understood more by their dictionary meaning than by their linkage to the theory, methodology, or field they belong to. This presents a serious challenge for machine translation, which often finds it impossible to consider the wider cognitive context; instead, it renders decisions on translation based on statistical or predictive patterns. Nevertheless, the scientific research community and the academic sphere at large seek to hasten the transfer of knowledge and bridge the linguistic gaps existing between academic communities. It is this that denotes the need for translation tools to be executed swiftly and accurately. In fact, the use of

machine translation in this regard is itself a subject of intense debate, wherein arguments concerning translation quality and reliability come straight into focus as they pertain to the faithful and safe transmission of scientific information, especially in sensitive areas such as health, medicine, and bioengineering. From this background, the study aims at shedding light on machine translation from an exact scientific perspective. Specifically, it attempts to determine to which degree the system performs well or poorly in translating scientific texts, especially as it seeks to preserve terminological accuracy versus contextualization. The study also addresses the most pressing challenges faced by these systems and furnishes scientific advice that would advance their improvement in the field. By analyzing and classifying translations in various fields according to a number of criteria that provide the most recurring erroneous translation patterns, such as the incorrect handling of technical terms, cases of literal translation, or cases of syntactic ambiguity, this research contributes to a better understanding of the present limitations of and potentials for machine translation. The study further provides informed suggestions for improving the quality of translations so as to assist the larger aim of making scientific knowledge more widely accessible worldwide in the digital era.

With the rapidity of technological development, machine translation is a necessity to bridge gaps among languages, particularly in scientific literature, which is marked by intricacy, nuance, and terminological specificity. From early rule-based systems to today's neural machine translation algorithms such as those used in Google Translate and DeepL, the history of this technology is a progression from literal translation to sophisticated, contextual systems powered by artificial intelligence. While now used generally in personal, educational, and professional applications, use in technical scientific texts raises fundamental issues of accuracy, validity, and conceptual correctness. Scientific writing often weaves together intricate sentence grammar, specialized terminology, and heavy reliance on context—dimensions challenging even advanced translation software to handle. Even in this type of writing, the meanings of words are not merely determined by lexical meanings but also by how words fit into a

theoretical or methodological frame of reference. As a result, preserving conceptual meaning, coherence, and terminological consistency in translation has now been an uphill task, one that computer translation systems are yet to attain successfully.

As increasing amounts of machine translation are being utilized in order to facilitate scientific communication worldwide, the accuracy of the translated scientific data has real-world implications. Misunderstanding medical research studies, policy reports, or technical specifications can result in real-life effects like violations of ethics, miscommunication, or intellectual property loss. Research communities and researchers are thus under pressure to provide speed and accuracy in providing multilingual knowledge, particularly in such sensitive fields of health care, engineering, and environmental science.

Hence, the objective of the present study is to undertake a critical evaluation of the performance of machine translation systems in the translation of scientific texts and their ability to balance terminological accuracy with context. Putting focus on this subject helps in advancing machine translation while also enhancing the level of scientific documents in the world and offers a more profound insight into the interaction between language and technology in relation to the digital opportunities of the present time.

Statement of the problem :

Although machine translation technologies have made many advancements, the main issue remains their ability to create a balance in terminology correctness and context. Therefore, we ask the question, can machine translation create a reliable scientific translation without losing context ?

Significance of the Study :

The value of the study is based on a field where artificial intelligence meets scientific knowledge and is examining the machine's capacity to carry

out something a human would do, relying heavily on intuition and deep analysis.

Research Questions :

1. How accurate is machine translation in scientific texts ?
2. What extent does machine translation take into account the overall context of the text ?

Research Objectives :

- To analyze the performance of machine translation in scientific texts.
- To identify the challenges facing machine translation in maintaining accuracy and context.

Literature Review :

Machine translation has been examined in numerous studies, especially in relation to general and technical translation. Some studies have in fact focused on Machine Translation using deep learning and neural networks, without studying specific disciplines like scientific translation. For instance, the literature on Neural Machine Translation (NMT) describes how NMT enables a deeper understanding of the context of translated words compared to traditional statistical approaches, which ultimately leads to improvements in fluency and coherence over prior and current machine translation. NMT, however, has largely been studied with respect to general works; there is therefore a question about whether the methods of NMT express similarly powerful benefits in specialized works that demand domain knowledge and precise terminology. Other kinds of studies have concentrated urgent issues in technical document translation ; for example, translating user manuals, patents, and legal documents. As these studies suggest, technical translator researchers outline the need for terminological consistency, but also how problematic terms and specialized domain jargon create obstacles to consistency. Of course, scientific texts also contain similar problems; however, scientific studies uniquely struggle with the greater conceptual structures and exact contextualization of their pioneering works. In that regard, some researchers claim that machine translation

systems should be modified or trained using specialized corpora for scientific translation. Various recent studies have focused directly on the application of machine translation to biomedical and life sciences fields, acknowledging the serious implications of misinterpretations. These studies emphasize the necessity of incorporating ontologies and domain-specific glossaries in machine learning models for achieving terminological correctness. Studies on post-editing further indicate that human involvement is also necessary in correcting machine-translated texts, particularly to maintain scientific discourse integrity. Despite all these advances, there is a consensus that current machine translation programs, while helpful for draft or casual use, are not yet completely reliable for the finesse demands of scientific translation. Scientific terminology's complexity, coupled with its dependence on both linguistic and conceptual accuracy, calls for continued investigation into the development of more sophisticated, context-dependent models of translation. This research expands on this literature base by systematically assessing the performance of leading machine translation engines in translating scientific texts across varied fields and determining potential areas of difficulty and development.

Theoretical Framework:

This research focuses almost exclusively on contemporary translation theories that embrace the idea of context as central to the translation process. This work is based largely on Firth's (1957) context theory, which says, "you shall know a word by the company it keeps." This theory implies that all meaning is dependent on context, both linguistic context and situational context. The theory implies a rejection of literal or word-for-word translation ; it says that translation must not just be an understanding of words, but of the world's semantic environment in which terms are present. This is important in scientific translation because there are sophisticated meanings of terminology, which are situated within larger conceptual frameworks and methodological frameworks. With Firth's context theory applied to scientific translation, I can explain here why context, whatever that may be, is difficulty in keeping accuracy and integrity. Besides Firth's theory, regarding translation, I am referring to

Skopos theory with an emphasis on context, which emphasizes the "purpose" (or skopos) of said text. Skopos theory means the way the translator translates with and on behalf of the function and intended audience. In science texts, the purpose is sharing information with the correct terminology; the purpose for the translation is to document, make clear, and communicate, because the audience may or may not familiar with those ideas or concepts. When translating, there needs to be some understanding of how terminology is used within context and intention, and that has to be understood and maintained in communication to the reader. Skopos theory thus supports the view that translation is not merely linguistic transfer but a purposeful act aimed at fulfilling specific communicative goals.

Related Studies :

Research on machine translation has evolved from rule-based and statistical models to advanced neural machine translation (NMT) systems, which have improved fluency and context handling. However, these systems still struggle with rare and domain-specific scientific terminology. Studies focusing on scientific texts, especially in biomedical fields, highlight frequent errors in terminology and context that can lead to misleading translations. Human post-editing remains essential to ensure accuracy and preserve the intended meaning, particularly in high-stakes disciplines. Comparative evaluations of popular tools like Google Translate and DeepL show that while useful for initial drafts, they often require significant human revision to maintain scientific precision. Overall, the literature agrees that machine translation alone is not yet fully reliable for specialized scientific translation, underscoring the need for ongoing improvements and human-machine collaboration.

Methodology :

The study relied on a descriptive-analytical approach, with comparative applications between machine and human translations of authentic scientific texts taken from peer-reviewed academic articles.

Data Collection :

Data was collected from scientific sources published in English, translated using machine translation tools, and then compared with human translations.

Data Analysis :

A qualitative analysis was conducted of translations of scientific phrases, focusing on conveying scientific terminology and context, and identifying common errors.

Study Sample :

The sample included 10 scientific texts from various fields, including biomedicine, applied physics, and artificial intelligence.

Content Analysis :

Content analysis tools were used to classify error types, such as omission of context, literal translation, or misuse of terminology.

Study Instruments :

The study tools included Google Translate, DeepL, and a set of validated human translations, in addition to translation quality assessment models (such as BLEU and METEOR).

Results and Discussion :

The results showed that machine translation offers acceptable performance in terms of terminological accuracy, but it suffers from significant problems in conveying precise context, especially in texts requiring in-depth specialized understanding. Look at the following table.

Table.1				
No.	**Original Scientific Text (Excerpt)**	**Machine Translation (Google Translate / DeepL)**	**Common Errors Identified**	**Results Summary**
1	Biomedicine : "The activation of T-cells plays a crucial role in the immune response to viral infections."	Google : "تنشيط الخلايا التائية يلعب دورًا حاسمًا في الاستجابة المناعية للعدوى الفيروسية ". DeepL : "يعد تنشيط الخلايا التائية أمرًا بالغ الأهمية في استجابة الجهاز المناعي للعدوى الفيروسية".	Literal translation of “activation” Minor omission of immunological nuance	DeepL better preserved meaning; both tools simplified complex immunological terms.
2	Applied Physics: "Quantum entanglement enables instantaneous state correlation between particles regardless of distance."	Google : "التشابك الكمومي يتيح الترابط الفوري للحالة بين الجسيمات بغض النظر عن المسافة ". DeepL : "يُمكّن التشابك الكمومي الارتباط اللحظي بين حالات الجسيمات بغض النظر عن المسافة".	Minor term consistency issues Google Translate less formal phrasing	Both tools accurate with terminology; DeepL more coherent scientifically.
3	Artificial Intelligence : "Deep learning models require	Google : "تتطلب نماذج التعلم العميق مجموعات بيانات كبيرة للتعميم	Minor sentence flow differences No significant terminology errors	Both tools accurate; human translations preferred formal scientific style.

	large datasets to generalize effectively across different tasks."	بشكل فعال عبر مهام مختلفة ". DeepL : "تحتاج نماذج التعلم العميق إلى مجموعات بيانات كبيرة لكي تعمم بشكل فعال عبر مهام مختلفة".		
4	Environmental Science : "Climate change accelerates the melting of polar ice caps, contributing to sea-level rise."	Google : "يزيد تغير المناخ من ذوبان الأغطية الجليدية القطبية، مما يساهم في ارتفاع مستوى سطح البحر ". DeepL : "يسرع تغير المناخ ذوبان الأغطية الجليدية القطبية، مما يؤدي إلى ارتفاع مستوى سطح البحر".	Slight difference in verb choice affecting tone Literal phrase usage	DeepL captures urgency better; Google is more literal but clear.
5	Genetics : "Gene editing technologies like CRISPR-Cas9 have revolutionized molecular biology research."	Google : "لقد أحدثت تقنيات تحرير الجينات مثل CRISPR-Cas9 ثورة في أبحاث البيولوجيا الجزيئية ". DeepL : "غيّرت تقنيات تحرير الجينات مثل CRISPR-Cas9 بشكل جذري أبحاث البيولوجيا الجزيئية".	Terminology mostly accurate Slight difference in formality and style	Both tools performed well; human translations still better in tone consistency.

6	Materials Science: "Nanomaterials exhibit unique properties that differ significantly from bulk materials."	Google : "تظهر المواد النانوية خصائص فريدة تختلف بشكل كبير عن المواد الكبيرة ". DeepL : "تتميز المواد النانوية بخصائص فريدة تختلف كثيرًا عن المواد الضخمة".	Misinterpretation of "bulk materials" as "large materials" in Google Translate	DeepL better grasped scientific nuance ; Google slightly inaccurate in terminology.
7	Neuroscience : "Synaptic plasticity is fundamental to learning and memory formation in the brain."	Google : "اللدونة المشبكية أساسية للتعلم وتشكيل الذاكرة في الدماغ ". DeepL : " اللدونة المشبكية هي أساس التعلم وتكوين الذاكرة في الدماغ".	Minor syntax differences Both correct terminology	Both tools accurate; DeepL more natural phrasing.
8	Pharmacology: "Drug metabolism in the liver involves cytochrome P450 enzymes which influence drug efficacy."	Google : "يتضمن استقلاب الدواء في الكبد إنزيمات السيتوكروم P450 التي تؤثر على فعالية الدواء ". DeepL : "يشمل استقلاب الأدوية في الكبد إنزيمات السيتوكروم P450 التي تؤثر على فعالية الدواء".	Minor pluralization difference Terminology consistent	Both tools precise; slight stylistic preference for DeepL.
9	Astrophysics: "Black holes are	Google : "الثقوب السوداء هي	Minor stylistic differences Both	Both tools delivered high-quality scientific

	regions of spacetime exhibiting gravitational acceleration so strong that nothing can escape."	مناطق في الزمكان تظهر تسارعًا جاذبيًا قويًا جدًا بحيث لا يمكن لأي شيء الهروب ". Google : "الثقوب السوداء هي مناطق من الزمكان تتميز بتسارع جاذبي قوي جدًا بحيث لا يستطيع أي شيء الهروب".	accurate terminology	translations.
10	Chemistry : "Catalysts accelerate chemical reactions without being consumed in the process."	Google : " تسرّع المحفزات التفاعلات الكيميائية دون أن تُستهلك في العملية ". DeepL : "تسرّع المحفزات التفاعلات الكيميائية دون أن تستهلك خلال العملية".	Minor difference in verb negation phrasing	Both tools accurate; DeepL slightly more formal.

Example 1 : Biomedicine

English : Activation of T-cells is one of the critical processes positioned in the immune response against viral infections.

Google : تنشيط الخلايا التائية يلعب دورًا حاسمًا في الاستجابة المناعية للعدوى الفيروسية

DeepL : يعد تنشيط الخلايا التائية أمرًا بالغ الأهمية في استجابة الجهاز المناعي للعدوى الفيروسية

Problem : Google translated the text more literally. DeepL went for the flow and clarity.

Result : DeepL maintained the scientific nuance and tone better.

Example 2 : Applied Physics

English : Quantum entanglement allows instantaneous correlation of state between particles irrespective of distance.

Google : التشابك الكمومي يتيح الترابط الفوري للحالة بين الجسيمات بغض النظر عن المسافة

DeepL : يُمكّن التشابك الكمومي الارتباط اللحظي بين حالات الجسيمات بغض النظر عن المسافة

Problem : Google used the singular form "الحالة" whereas DeepL rightly used the plural "حالات" to suit the context.

Result : DeepL was more technically accurate; Google was understandable.

Example 3 : Artificial Intelligence

English : Deep learning models require large datasets to generalize effectively across different tasks.

Google : تتطلب نماذج التعلم العميق مجموعات بيانات كبيرة للتعميم بشكل فعال عبر مهام مختلفة

DeepL: تحتاج نماذج التعلم العميق إلى مجموعات كبيرة من البيانات كي تعمم بفعالية في المهام المختلفة

Problem : Minor difference in flow and style ; the terminology remained accurate.

Result : Both translations were reliable, with DeepL providing slightly more naturally flowing phrasing.

Example 4 : Environmental Science

English : Climate change accelerates the melting of polar ice caps, contributing to sea-level rise.

Google : يزيد تغير المناخ من ذوبان الأغطية الجليدية القطبية، مما يساهم في ارتفاع مستوى سطح البحر

DeepL : يسرع تغير المناخ ذوبان الأغطية الجليدية القطبية، مما يؤدي إلى ارتفاع مستوى سطح البحر

Problem : "يزيد" in Google is more neutral, while "يسرع" in DeepL better conveys urgency.

Result : DeepL captured the intended tone and urgency more effectively.

Example 5 : Materials Science

English : Nanomaterials exhibit unique properties that differ significantly from bulk materials.

Google : تظهر المواد النانوية خصائص فريدة تختلف بشكل كبير عن المواد الكبيرة

DeepL : تتميز المواد النانوية بخصائص فريدة تختلف كثيرًا عن المواد الضخمة

Problem : Google incorrectly translated "bulk materials" as "large materials".

Result : DeepL provided a more accurate scientific translation.

The most common errors are: literal translations, small terminological inconsistencies, or omission of delicate context.

Google Translate: Generally good meaning but uses scientific language less formally or precisely at times.

DeepL: More relative, better at recognizing scientific terminology and phrasing.

Human Translations: Most accurate, especially regarding complex conceptual fidelity and contextual fidelity.

Quality scores showed that in analyses using BLEU and METEOR that DeepL's scores were 10-15% higher on average. Look at the following table:

Table.2

Error Type	Average % Error Rate
Omission of Context	12%
Literal Translation	18%
Misuse of Terminology	14%

Error Type : Omission of Context

Average Percentage of Error : 12 percent. Impact on BLEU and METEOR: Such omission of context affects the alignment of sentences and recall, particularly regarding the sentences with some conceptual or rather abstract language.

Example from Neuroscience :

English : Synaptic plasticity is fundamental to learning and memory formation in the brain.

Google : اللدونة المشبكية أساسية للتعلم وتشكيل الذاكرة في الدماغ

DeepL : اللدونة المشبكية هي أساس التعلم وتكوين الذاكرة في الدماغ

Problem : The Copula ("is") was dropped in the Google translation, which makes the sentence lose clarity.

Effect on the scores : BLEU penalizes the absence of such structural normalization. Similarly, METEOR punishes the partial match and lack of context.

Result : DeepL fares better because its translation retains structure and context.

Error Type : Literal Translation

Average Error Rate : 18 percent. Impact on BLEU and METEOR : Literal uses of words frequently have different meanings in the reference translation, resulting in a lower overall accuracy and readability.

Example from Materials Science :

English : Nanomaterials have unique properties that are quite different from bulk materials.

Google : تظهر المواد النانوية خصائص فريدة تختلف بشكل كبير عن المواد الكبيرة

DeepL : تتميز المواد النانوية بخصائص فريدة تختلف كثيرًا عن المواد الضخمة

Problem : Google incorrectly translates "bulk materials" as "large materials"

Impact : BLEU drops because of lack of vocabulary match. METEOR will penalize the incorrect word and loss of scientific accuracy.

Result : DeepL used the proper term and therefore will get a better score.

Error Type : Misuse of Terminology

Average Error Rate : 14 percent. BLEU and METEOR

Impact : Inaccurate use of scientific terms reduces both exact match and semantic accuracy.

Applied Physics Example :

English : Quantum entanglement facilitates instantaneous state correlation between particles regardless of distance.

Google : التشابك الكمومي يتيح الحالة الفورية للتعلق بين الجسيمات بغض النظر عن المسافة

DeepL : يُمكّن التشابك الكمومي المشاركة اللحظية بين حالات الجسيمات بغض النظر عن المسافة

Problem : Google uses the singular "state" instead of the correct plural "states," sacrificing technical accuracy.

Impact : BLEU inflicts penalty on wrong form. METEOR decreases score because of semantic error in a scientific term.

Result : DeepL is more accurate in scientific phrasing and has a better score.

Discussion :

The evaluation of machine translations of ten scientific texts from various content areas identified several challenges in accurately representing scientific content with forms of machine translation, particularly as it relates to context, specificity of terms, and also with verbatim translations. Omission of Context was a notable error category averaging about 12% across the corpus. This number is important because it reflects a consistent limitation with many machine translations, which often struggle to consider the broader conceptual context or framework for scientific terms and phrases. Scientific texts are not simply about word-for-word replacement,

they require an understanding of theories and methodologies associated with that content. The absence of consideration of context available to a human translator with a scientific background creates machine translations that are grammatically correct. However, they are often missing substantive considerations required for a knowledge transfer. Our analysis of machine translation of scientific texts raised several important issues, particularly relating to omission of context, literal translation, and misuse of terminology. Although literal translations represent the smallest number of errors at 18 percent, they may have unintended meanings that have specific finding in a domain, and thus unintentional negative implications and miscommunication. For example, "bulk materials" was translated as "large materials." Misuse of terminology accounted for an average of 14 percent of errors, but was more nuanced - again, may have far-reaching implications, scubas terminology in applied physics, biomedicine and nonsense in the formal (i.e., publication and grant writing) and collaboration context. DeepL outperformed Google Translate in every way, particularly for maintaining scientific tone, phrasing, and according to our raters, natural Arabic syntax. This translates into - on average 10 to 15 percent lead in BLEU and METEOR score. While these tools are head-and-shoulders better than Google Translate, they still fall short of a human translator for implicit meanings, appropriate equivalents in context, or coherence. Although, clearly, the human translator has subject-matter knowledge and conceptual fluency that machines lack. For these reasons, these developments in machine translation, such as DeepL, could be beneficial for a draft for informal use and for understanding, but they should not be used for purposes that that impact direct scientific communication, whether risk or opportunity. The future of machine translation will ultimately depend on developing services that are integrated with domain knowledge, modeling context to support accuracy, and adaptive learning for specialized development.

Recommendations :

1. Increase contextual understanding in machine translation models
2. Incorporate specific glossary of scientific terminology databases
3. Use new approaches to utilize both machine and human translation

Conclusion :

The results of the study conclude that machine translation is only a part of scientific publishing and is not enough by itself to translate exact science texts that require intensive contextual understanding. In the end, the two machine translation tools examined in this study, DeepL was consistently better than Google Translate. DeepL produced translations with more contextual meaning and scientific detail. Google Translate was and continues to be freely available and was improving in its accuracy, but in comparison, produced translated sentences that had more literal translations, more terminological inaccuracies, and more generic scientific texts. Both tools ultimately fall short of human translation quality, especially in strict technical and sensitive scientific areas. This is an interesting finding for the scientific realm in terms of the implications of using machine translation. If reliable machine translation can be achieved, it could have considerable merit to the rapid sharing of knowledge internationally between scientists, by bridging language and cultural barriers and increasing collaboration opportunities. But as with anything else, the limitations of the machine translation quality need to acknowledged as we move forward, particularly in sensitive areas where mistranslation could produce disastrous results such as medicine, bio-engineering and environmental science. Future work will need to enhance the context-level understanding of machines translation systems through the use of the specific domain datasets, the necessary feedback of experts, and generic learnings. The combination of machine translation or models with human post-edits could also be a reasonable compromise that ensures efficiency in their uptake whilst ensuring accuracy.

References

Alvarez, S., Oliver, A., & Badia, T. (2020). Quantitative analysis of post-editing effort indicators for NMT. Proceedings of the 22nd Annual Conference of the European Association for Machine Translation (EAMT).[https://aclanthology.org/2020.eamt-1.44/] (https://aclanthology.org/2020.eamt-1.44/).

Bahdanau, D., Cho, K., & Bengio, Y. (2016). Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473. [https://arxiv.org/abs/1409.0473] (https://arxiv.org/abs/1409.0473).

Gaspari, F., Toral, A., Naskar, S., & Way, A. (2014). Online MT services and professional translators. Perspectives: Studies in Translatology, 22 (4), 497–515.[https://doi.org/10.1080/0907676X.2014.948018] (https://doi.org/10.1080/0907676X.2014.948018).

Guerreiro, N. M., Alves, D., Waldendorf, J., Haddow, B., Birch, A., Colombo, P., & Martins, A. F. T. (2023). Hallucinations in large multilingual translation models. arXiv preprint arXiv: 2303.16104. [https://arxiv.org/abs/2303.16104]

Koehn, P. (2020). Neural machine translation. Cambridge University Press.

Lee, S., Lee, J., Moon, H., Park, C., Seo, J., Eo, S., Koo, S., & Lim, H. (2023). A survey on evaluation metrics for machine translation. Mathematics, 11(4), 1006.[https://doi.org/10.3390/math11041006] (https://doi.org/10.3390/math11041006)

Li, F., Chen, J., & Zhang, X. (2023). A survey of non-autoregressive neural machine translation. Electronics, 12(13), 2980. [https://doi.org/10.3390/electronics12132980] (https://doi.org/10.3390/electronics12132980)

Niño, A. (2009). Machine translation in foreign language learning: Language learners' and tutors' perceptions. ReCALL, 21(2), 241–258. [https://doi.org/10.1017/S0958344009000172]

O'Brien, S. (2005). Methodologies for measuring the correlations between post-editing effort and machine translatability. Machine Translation, 19(4), 227–248.[https://doi.org/10.1007/s10590-006-8780-3] (https://doi.org/10.1007/s10590-006-8780-3)

Popović, M. (2020). Error classification and analysis for machine translation output. Computational Linguistics, 45 (4), 603–632. https://doi.org/10.1162/coli_a_00383

Sarti, G., Bisazza, A., Guerberof Arenas, A., & Toral, A. (2022). DivEMT: Neural machine translation post-editing effort across typologically diverse languages. arXiv preprint arXiv:2205.12215. [https://arxiv.org/abs/2205.12215]

Tiedemann, J. (2011). Bitext alignment. Synthesis Lectures on Human Language Technologies,4(2),1–165. [https://doi.org/10.2200/S00371ED1V01Y201108HLT013] (https://doi.org/10.2200/S00371ED1V01Y201108HLT013)

Vasconcellos, M., & León, S. (1985). SPANAM and ENGSPAN: Machine translation at the Pan American Health Organization. In Computers and Translation: A Translator's Guide (pp. 33–56). John Benjamins.[https://doi.org/10.1075/z.20.07vas] (https://doi.org/10.1075/z.20.07vas)

Xiao, Y., Wu, L., Guo, J., Li, J., Zhang, M., Qin, T., & Liu, T.-Y. (2022). A survey on non-autoregressive generation for neural machine translation and beyond. arXiv preprint arXiv:2204.09269. [https://arxiv.org/abs/2204.09269] (https://arxiv.org/abs/2204.09269)

Zaretskaya, A., Vela, M., Corpas Pastor, G., & Seghiri, M. (2016). Measuring post-editing time and effort for different types of machine translation errors. New Voices in Translation Studies. [https://doi.org/10.14456/nvts.2016.21] (https://doi.org/10.14456/nvts.2016.21)