Using Translation Techniques to Characterize MT Outputs

Sergi Alvarez-Vidal, Maria de Campo, Christian Olalla-Soler, Pilar Sánchez-Gijón
Universitat Autònoma de Barcelona (UAB)
{sergi.alvarez, maria.docampo, christian.olalla, pilar.sanchez.gijon}@uab.cat

Abstract

While current neural machine translation (NMT) and generative pre-trained transformer (GPT) models improve fluency and context awareness, they struggle with creative texts, where figurative language and stylistic choices are crucial. Current evaluation methods fail to capture these nuances, which require a more descriptive approach. We propose a taxonomy based on translation techniques to assess machine-generated translations more comprehensively. The pilot study we conducted comparing human and machine-produced translations reveals that human translations employ a wider range of techniques, enhancing naturalness and cultural adaptation. NMT and GPT models, even with prompting, tend to simplify content and introduce accuracy errors. Our findings highlight the need for refined frameworks that consider stylistic and contextual accuracy, ultimately bridging the gap between human and machine translation performance.

1 Introduction

Rapid advancements in neural machine translation (NMT) and generative pre-trained transformer (GPT) models have significantly improved the quality of machine-generated translations in recent years. In many cases, these models achieve an output product that closely resembles human translation (Jiao et al., 2023; Wang et al., 2023), making it increasingly difficult to describe or evaluate their performance using traditional metrics. Although early claims suggested that machine translation (MT) had reached parity with human translation (Hassan et al., 2018), subsequent studies have challenged these assertions, underscoring the persistent difficulties of evaluating machine-generated output in a way that captures their full complexity (Toral et al., 2018; Läubli et al., 2020). These debates further emphasize the limitations of current assessment methods, particularly their inability to account for the contextual and stylistic nuances (Wang et al., 2024) that professional translators consider essential.

One of the most pressing challenges in this context is the translation of creative texts, such as literature and marketing content. Unlike technical or informational texts, which often follow predictable structures and terminology, creative texts rely heavily on figurative language, including irony, metaphor, and ambiguous phrasing. These elements often lead to overly literal, word-for-word translations that do not convey the intended meaning in the machine-translated text (Guerberof-Arenas and Toral, 2020). Although current GPT models offer notable improvements by considering broader contextual relationships in sentences (Castilho et al., 2023), they still struggle with the complexities of creative expression.

Current evaluations are usually based on automatic metrics such as BLEU (Papineni et al., 2002) or COMET (Rei et al., 2020), or on manual evaluations that produce a list of errors and their severity, such as the MQM taxonomy (Lommel et al., 2014). However, these metrics mainly focus on the traditional accuracy and fluency paradigms, which do not account for any stylistic variation. Recent research has even shown that the inclusion of machine-translated texts in test data can significantly affect the results of evaluation outcomes. For example, Graham et al. (2020) found that MT systems may appear to perform better or worse depending on the nature of the test data.

Thus, we need to explore alternative approaches to describe and assess these texts in accordance with the contexts in which they are intended to be used. In contexts where both the conveyed information and the expressive or persuasive function of the text are essential, human translators frequently

Proceedings of Machine Translation Summit XX Volume 1, pages 619–627, June 23-27, 2025

employ a range of techniques to help the target audience grasp the subtle nuances of the original text. These strategies ensure that not only the content, but also the intended impact of the text is effectively conveyed in the translated version. If texts translated using NMT and GPT models are employed in the same scenarios where human translators apply these techniques, it is worth considering whether these techniques can also serve to describe and, consequently, evaluate the quality of machine-generated translations.

In this context, Translation Studies provide a rich theoretical framework that can offer more nuanced descriptive criteria. Specifically, we develop a taxonomy partially based on the translation techniques defined by Molina and Hurtado (2002). Their framework categorizes the translation techniques employed by human translators, which can serve as a benchmark for describing machine translations at a deeper level.

Using translation techniques such as modulation, amplification and explicitation, our proposed method aims to capture the complexity of translation beyond literal equivalence, helping us describe machine translation outputs or *machine translationese*. This approach enables us to assess how well MT models handle pragmatic and linguistic challenges, including idiomatic expressions, register changes, and cultural adaptation, thus providing a more comprehensive understanding of their strengths and weaknesses.

The remainder of this paper is structured as follows. Section 2 reviews related work on *translationese* regarding MT, highlighting some of the key concepts of its characterization. Section 3 introduces the proposed framework based on translation techniques and its theoretical underpinnings. Section 4 presents the setup and methodology used to conduct a pilot study of this framework, followed by the results in Section 5. Finally, Section 6 concludes the paper and outlines future research directions.

2 Human Translation vs Machine Translation

The study of differences between translated texts and non-translated texts has long been a central focus of translation studies research, with early research identifying distinct linguistic features that describe what has been called *translationese*. Toury 2012 differentiates between the law of interference, which refers to the elements of the source text that are retained in the translation, and the law of growing standardization, which relates to the tendency to apply the norms of the target language and culture to the translation product. Thus, any final translation is the hybrid result of the application of both laws.

Chesterman (2004) makes a distinction between S-universals and T-universals. S-universals are features that can be traced back to the source text. T-universals, on the other hand, are features that should be studied by comparing translated texts to non-translated texts in the target language, using a comparable corpus. They include features such as simplification, untypical patterning, and underrepresentation of target-language-specific items.

Baker (1993) suggests there are several translation universals, which are linguistic features that tend to characterize translated texts regardless of the language pair or direction of translation. These include simplification, where translations exhibit reduced structural and lexical complexity; explicitation, the tendency to render implicit information more explicit; normalization, which aligns translations more closely with conventional target language norms; leveling-out, which results in reduced variation across different text types; and interference, where source language structures influence the target text.

Corpora have been used extensively to study *translationese*. For example, Corpas Pastor (2008) argues that translated texts include lower lexical diversity, shorter sentence structures, and increased explicitation. These tendencies emerge due to the translator's dual commitment to preserving source meaning while ensuring readability in the target language. Empirical studies using comparable corpora have consistently shown that *translationese* manifests across languages, regardless of the specific translation directions (Volansky et al., 2015).

Human translations and machine translations have also shown divergences at the morphosyntactic level. Luo et al. (2024) conduct a large-scale fine-grained comparative analysis across three language pairs and show MT is consistently more conservative than human translations, as it shows less morphosyntactic diversity, more convergent patterns, and more one-to-one alignments.

As MT technology advances, researchers have begun to investigate whether similar patterns can be detected in MT-generated texts and post-edited (PE) translations (Castilho and Resende, 2022;

Toral et al., 2018). Some studies suggest that PE texts inherit certain traits from raw MT output, such as reduced lexical diversity and terminological consistency that align more closely with machine-generated texts than with human translations. For instance, Vanmassenhove et al. (2021) identify a loss in lexical richness in MT output, which could subsequently influence the characteristics of post-edited texts. Toral (2019) finds that post-edited documents have lower lexical variety and lower lexical density than human translations. Moreover, sentence length and parts-of-speech in post-edited texts are more similar to the source language than those in human translations.

A study by Zhu et al. (2024) examines translation relations to identify differences between NMT and human translations. The findings reveal that NMT systems tend to rely more heavily on literal translations compared to human translators, especially in the use of semantic-level translation techniques. The advent of large language models (LLMs) and GPTs has introduced the concept of *generatese*, referring to the distinct linguistic patterns produced by these models during text generation tasks, including translation.

He et al. (2024) investigate whether LLMs can mimic human translation strategies by analyzing source sentences and inducing translation-related knowledge such as keywords and topics. Their research shows that while LLMs can exhibit human-like translation strategies, there are challenges to reducing errors such as hallucinations and mistranslations, which are often associated with *generatese*.

Comparative analyses between human translations and machine-generated texts have highlighted notable differences. A study by Chen et al. (2024) proposes an iterative prompting approach for LLMs to self-correct translations. Interestingly, while this method reduces string-based metric scores, neural metrics suggest comparable or improved quality. These refined translations achieve better fluency, although other challenges related to *generatese* still remain. Other studies also suggest that LLMs generate translations that deviate more from the source text than those produced by NMT models (Vilar et al., 2023; Raunak et al., 2023).

3 Framework of Translation Techniques

Translation techniques play a fundamental role in Translation Studies, serving as essential tools to analyze and understand the procedures by which translators achieve equivalence between source and target texts, and have long been studied by translation scholars (Vinay and Darbelnet, 1958; Newmark, 1981, 1988; Chuquet and Paillard, 1989; Molina and Albir, 2002; Gibová, 2012). These techniques provide a framework for systematically identifying and categorizing the choices translators make during the translation process to address linguistic, cultural, and contextual challenges. Their significance extends to various aspects of translation theory and practice, contributing to improving translation quality and the development of pedagogical approaches.

Translation techniques allow for a structured approach to evaluating translation choices by offering a set of predefined categories that describe how equivalence is achieved at the micro-textual level. This systematic analysis helps identify patterns in translator behavior, and to compare different translations of the same text. By distinguishing techniques, we can better understand how translators navigate linguistic and cultural differences.

However, there is no consensus in academia on the classification and nomenclature of translation techniques. Vinay and Darbelnet (1958) were the first to publish a classification of translation techniques with a clear methodological purpose. They defined seven basic procedures operating on three levels of style and classified them between literal and oblique.

Nida (1964) suggests three types of translation techniques: additions, subtractions and alterations. These techniques are used to adjust the form of the message to the characteristic structure of the target language, to produce semantically equivalent structures, to generate adequate stylistic equivalences, and to produce an equivalent communicative effect.

Newmark (1988) uses the term *procedures* to classify translation techniques proposed by comparative linguists. These include: recognized translation, where an already accepted term is used even if it is not the most precise; functional equivalence, which replaces a term with a culturally neutral expression plus a qualifier; and naturalization, which adapts a source language word to the phonetic and morphological norms of the target language. He also introduces translation labels for provisional translations, often literal in nature. Additionally, Newmark allows for combining multiple procedures (doubles, triples, etc.) and includes synonymy as a separate category.

Molina and Hurtado (2002) modify and expand previous classifications. They isolate the concept of technique by focusing on the notion of functionality, situating it in relation to the text and the context. For our framework, we take into account previous research on MT-generated content (Sanchez-Gijón, 2024; Zhai et al., 2024) and we make an effort to group the different phenomena in order to simplify corpus annotation. We simplify the original set of 18 translation techniques and add *naturalness*, which should be understood as a habitual use of the language, free of grammatical errors, fluid in style, and without expressions that are strongly influenced by other languages (do Campo Bayón and Sánchez-Gijón, 2024). Below we define the translation techniques and illustrate them with some examples from the annotated segments of the pilot study detailed in Section 4, for the Catalan-English language pair:

- **Non-literal linguistic choices in the pursuit of naturalness** This technique involves a departure from the original text, showcasing creativity in form while maintaining the original content. The translator prioritizes fluency and idiomatic expression in the target language to achieve a natural-sounding result.

  CA: Ja devia tenir un senyal vermell a la cintura, però així que el vent m'havia sortit per la boca la cinta tornava a fer-me el martiri. [I must have already had a red mark on my waist, but as soon as the wind had left my mouth, the ribbon went back to tormenting me.]
  EN: I pictured the red weal round my waist, but the moment I started rushing and getting out of breath, the elastic sliced into me again.

- **Established equivalent** This refers to the use of pre-existing, widely accepted equivalents in the target language, such as titles of movies, books, or brand names. By opting for the established equivalent, the translator ensures coherence and consistency with conventional usage.

  CA: La meva reina, va dir [My queen, he said.]
  EN: He said, my darling.

- **Simplification** Simplification entails the reduction of information without omitting essential meaning. It includes generalization and linguistic compression, conveying the same message with fewer details. Example:

  CA: La cinta de goma a la cintura estrenyent, estrenyent (...) [The rubber band around my waist, tightening, tightening.]
  EN: The elastic cutting deep into my waist (...).

- **Omission** The omission technique involves deliberately leaving out specific information that may not be essential for the overall message. The resulting text remains functional and coherent despite the absence of the omitted element.

  CA: (...) i a cada banda de la cara la medalleta de l'orella. [and on each side of the face, the little medal on the ear.]
  EN: (...) and little medal-like ears.

- **Explicitation** This technique makes implicit details (whether linguistic or thematic) explicit in the target text. It can include clarifying pronouns based on the level of formality or providing additional gender markers. Example:

  CA: Tan petita i ja té promès? [so young and you already have a fiancé?]
  EN: 'Aren't you too young to have a fiancé?'

- **Amplification** Amplification involves adding or making explicit details that the original audience might infer naturally. This technique is particularly useful when cultural or contextual knowledge cannot be assumed in the target audience.

  CA: (...) i vinga riure [and he kept on laughing]
  EN: (...) and he laughed till he cried.

- **Adaptation** Adaptation consists of finding an equivalent expression in the target language and culture that serves a similar function, even if it is not an established term. This technique is central to the domestication strategy, making the text more accessible and relatable to the target audience.

  CA: (...) la meva mare morta i sense poder-me aconsellar [my mother dead and unable to advise me]
  EN: (...) my mother dead and gone and not around to give me advice

- **Fluency and accuracy errors** These errors occur when the translated text contains unnatural phrasing, awkward constructions, or

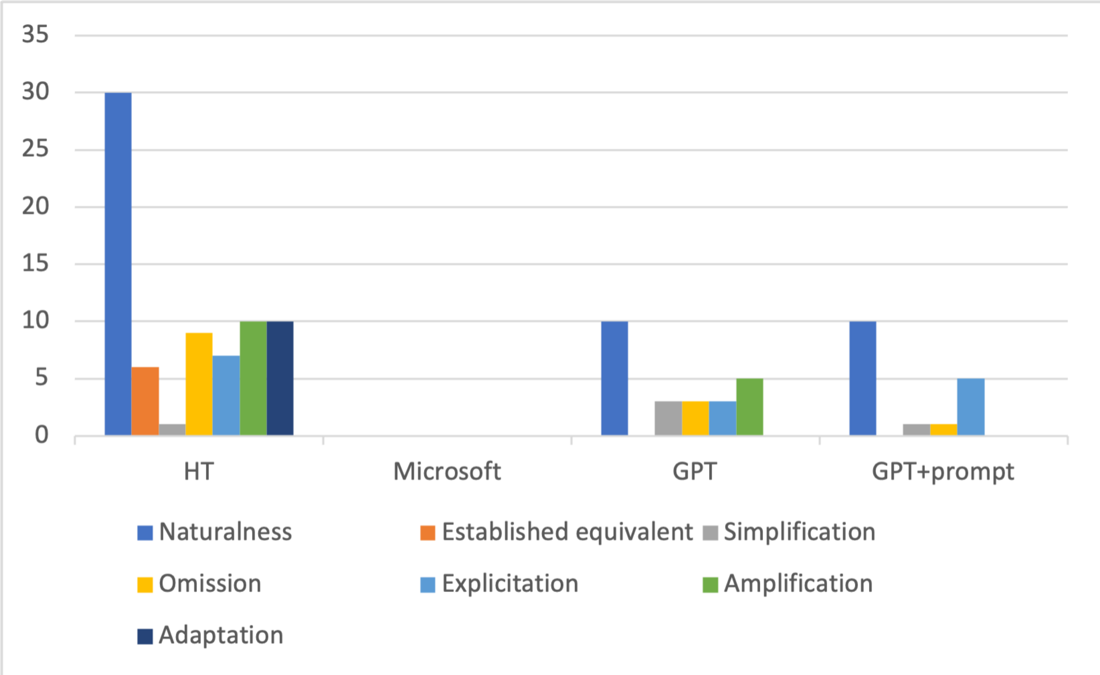

Figure 1: Use of techniques in the different translations

inaccuracies that may hinder comprehension or misrepresent the source text. They can include grammatical mistakes, stylistic inconsistencies, or mistranslations that affect the quality of the final output. These are the usual elements included in traditional evaluations and are incorporated in our annotation process to better understand the output translations in relation to usual evaluation techniques.

By applying these categories, we aim to gain deeper insights into the decision-making process of translators and the impact of various strategies on the final translated text.

4 Experimental setup and methodology

As an initial step following the selection of the translation techniques to be used for annotation, we decided to conduct a pilot study using one of the most renowned works in Catalan literature, *La Plaça del Diamant* by Mercè Rodoreda and its translation into English by Peter Bush in 2013. The novel was automatically segmented into sentences, and the first 60 segments were selected for the annotation process. We annotated the published translation into English and the translations produced by three MT engines. We used a NMT model (Microsoft Translator) and a GPT model (ChatGPT), as research shows these models translate broader contextual relationships across sentences better than NMT models (Castilho et al., 2023). Moreover, we used ChatGPT with a specific set of prompts to assess whether prompting techniques could improve the translation results for this type of text (Yamada, 2019; He, 2024).

We opted not to randomize the selection of segments, as the application of translation technique categories often relies on contextual references that extend beyond individual segments. Maintaining sequential order allowed us to preserve the coherence of the text and ensure that context-dependent techniques could be accurately identified and applied.

A relatively small number of segments was chosen for this pilot study, as its primary objective was twofold: first, to evaluate the relevance and applicability of the selected translation techniques; and second, to compare the results of the published human translation against raw machine translation (MT) outputs generated by NMT and GPT-based models with or without prompting techniques.

For each segment in the source language, four translations were annotated: (1) human translation, (2) Microsoft Translator translation, (3) ChatGPT translation without additional prompts, and (4) ChatGPT translation with specific prompts. For this version of ChatGPT, we introduced the following prompts in English, which described both the step-by-step actions followed by professional translators as well as some considerations regarding the

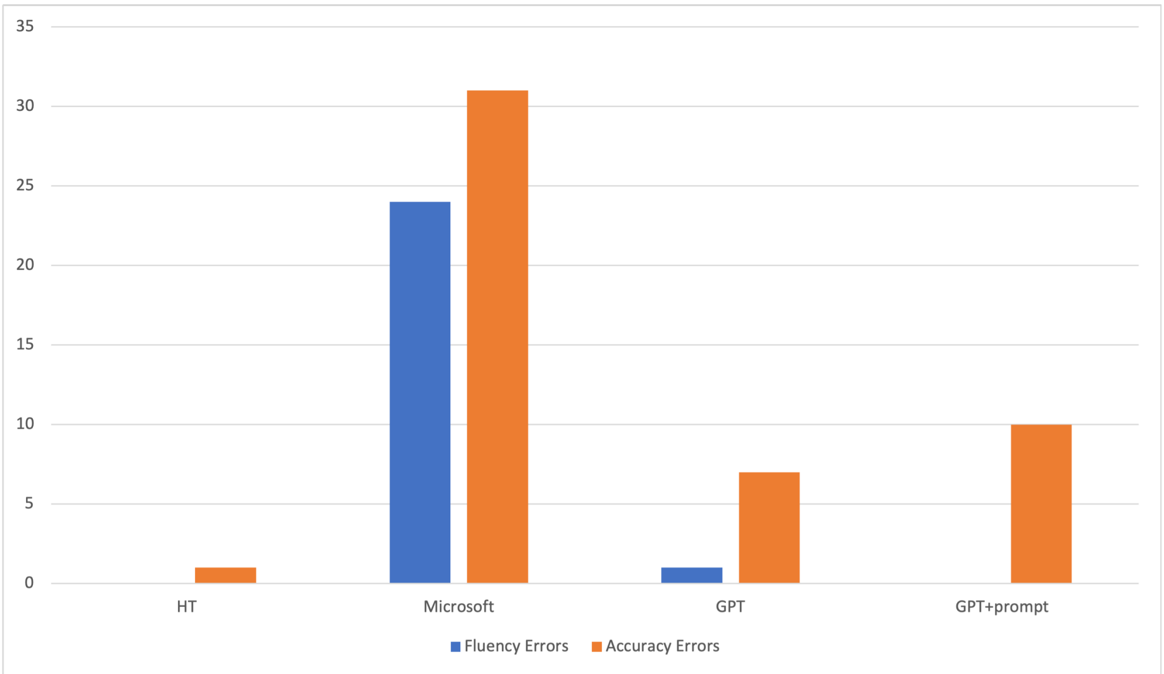

Figure 2: Accuracy and fluency errors

context and the text type:

Translate a literary text from Catalan (CA) to English (EN) while considering the cultural differences between the CA and EN readers. Follow a professional translator's strategy by considering the text's function, the cultural and social differences between the two audience groups, the author's style, and the text genre. Assume the EN reader is unfamiliar with CA culture, particularly regarding life in the 1950s in Barcelona and surrounding areas.

Steps to follow:

1. *Analyze the Original Text: Understand the text's purpose, the author's style, and specific cultural references unique to 1950s Barcelona.*
2. *Identify Cultural and Social Differences: Note key cultural elements that may need context or adaptation for an EN audience.*
3. *Translation Strategy: Adjust cultural references as needed to make them understandable without losing the text's authenticity. Maintain the original author's style and tone while ensuring it is accessible to an EN audience. Keep the genre conventions in mind to ensure the translated text aligns with expectations typical to that genre in English literature. Adapt for EN Readers: Provide additional context where necessary to enhance understanding of cultural nuances without altering the narrative.*
4. *Review and Revise: Ensure the final translation feels natural to an EN reader and accurately represents the original text's nuances.*

Output Format:

Provide the translated text in a natural and fluent English format, maintaining the original length as closely as possible while ensuring cultural clarity.

Each segment was annotated by two different annotators with previous experience in similar tasks. For the segments in which both annotators were not in agreement, a third annotator assessed the proposals and made a final decision.

5 Results

In Figure 1 we can see the results of the annotation process. Human translations include a higher number of translation techniques than any of the MT-produced translations, except for the simplification technique. In fact, this is one of the techniques that reduces source language information without any substitution or modification. However, human translation incorporates more omissions, which can be linked to the compensation process undertaken while translating, as in many other segments human translations incorporate techniques used to add more explicit information. It is also clear from the results that the NMT model (Microsoft) does not use any of the translation techniques and thus produces more literal translations.

From the annotated techniques, we can highlight the use of naturalness in the human translation, which is the most frequently applied. In the search to produce a text that engages the target reader and has the same impact as the source reader, the translator makes decisions that move away from word-to-word translation and incorporate a creative component. Moreover, human translations also include increased use of the adaptation of the content (for example, with names of people and places) and amplification of certain elements to highlight them in the translation.

In Figure 2 we can see the results for accuracy and fluency for all output translations. All outputs contain a considerable high number of inaccuracies or translations which do not convey the meaning of the source text. Once again, the NMT model produces the highest number of fluency and accuracy errors, which are highly reduced in the case of ChatGPT. An interesting result is that the inclusion of prompts increases the number of accuracy errors. This could be linked to the effort made by ChatGPT to create more literary and creative content when the instructions explicitly indicate it. The creation of this type of translations seems to have as a side-effect the increased number of hallucinations or errors in the translations.

6 Conclusion and Future Work

The improved quality of the translations produced by the NMT and GPT models makes it increasingly difficult to distinguish them from human-produced texts. Current evaluation metrics fail to account for the stylistic and contextual nuances that are crucial in human translation. The challenge is particularly evident in the translation of creative texts, where figurative language plays a key role in meaning-making.

To address these limitations, we proposed a framework based on translation techniques, inspired by established models in Translation Studies. Our pilot study comparing human, NMT and GPT-produced translations of *La Plaça del Diamant* reveals significant differences in translation strategies. Human translators employ a wider variety of techniques, such as amplification, naturalness, and adaptation, that contribute to more natural, culturally appropriate, and stylistically coherent translations. In contrast, NMT and GPT models, even with targeted prompts, tend to simplify content, favoring more literal renderings that sometimes fail to capture the expressive function of the source text. While prompting techniques can make GPT translations appear more creative, they also introduce a higher number of accuracy errors, suggesting a higher introduction of hallucinations.

These findings reinforce the need for refined evaluation frameworks that move beyond traditional metrics to incorporate a deeper analysis of textual adaptation and stylistic effectiveness. By systematically categorizing translation strategies, our approach provides a more comprehensive way to assess how well machine translations handle complex linguistic and cultural challenges. Future research should build on this framework by expanding corpus size, and exploring automated annotation methods to improve scalability. Ultimately, integrating translation techniques into MT evaluation can offer a more human-centric perspective, bridging the gap between computational advancements and the nuanced decision-making process of professional translators.

References

Mona Baker. 1993. Corpus Linguistics and Translation Studies — Implications and Applications. In Mona Baker, Gill Francis, and Elena Tognini-Bonelli, editors, *Text and Technology: In honour of John Sinclair*, page 233. John Benjamins Publishing Company.

Sheila Castilho, Clodagh Quinn Mallon, Rahel Meister, and Shengya Yue. 2023. Do online Machine Translation Systems Care for Context? What About a GPT Model? In *Proceedings of the 24th Annual Conference of the European Association for Machine Translation*, pages 393–417, Tampere, Finland. European Association for Machine Translation.

Sheila Castilho and Natália Resende. 2022. Post-Editese in Literary Translations. *Information*, 13(2).

Pinzhen Chen, Zhicheng Guo, Barry Haddow, and Kenneth Heafield. 2024. Iterative Translation Refinement with Large Language Models. In *Proceedings of the 25th Annual Conference of the European Association for Machine Translation (Volume 1)*, pages 181–190, Sheffield, UK. European Association for Machine Translation (EAMT).

Andrew Chesterman. 2004. Beyond the particular. In Anna Mauranen and Pekka Kujamäki, editors, *Translation Universals: Do they exist?*, Benjamins Translation Library, pages 33–49. John Benjamins Publishing Company.

Hélène Chuquet and Michel Paillard. 1989. *Approche* *linguistique des problèmes de traduction anglais-français*. Ophrys, Paris.

Gloria Corpas Pastor. 2008. *Investigar con corpus en traducción: los retos de un nuevo paradigma*, volume 49 of *Studien zur romanischen Sprachwissenschaft und interkulturellen Kommunikation*. Peter Lang.

María do Campo Bayón and Pilar Sánchez-Gijón. 2024. Evaluating NMT using the non-inferiority principle. *Natural Language Processing*, pages 1–20.

Katarína Gibová. 2012. Translation Procedures in the Non-literary and Literary Text Compared. *Bilingual Journal of Applied Linguistics*, 3(1):21–34.

Yvette Graham, Barry Haddow, and Philipp Koehn. 2020. Statistical Power and Translationese in Machine Translation Evaluation. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 72–81. Conference Name: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP) Place: Online Publisher: Association for Computational Linguistics.

Ana Guerberof-Arenas and Antonio Toral. 2020. The impact of post-editing and machine translation on creativity and reading experience. *Translation Spaces*, 9(2):255–282. Publisher: John Benjamins Publishing Company.

Hany Hassan, Anthony Aue, Chang Chen, Vishal Chowdhary, Jonathan Clark, Christian Federmann, Xuedong Huang, Marcin Junczys-Dowmunt, William Lewis, Mu Li, Shujie Liu, Tie-Yan Liu, Renqian Luo, Arul Menezes, Tao Qin, Frank Seide, Xu Tan, Fei Tian, Lijun Wu, Shuangzhi Wu, Yingce Xia, Dongdong Zhang, Zhirui Zhang, and Ming Zhou. 2018. Achieving Human Parity on Automatic Chinese to English News Translation. *arXiv preprint*. ArXiv:1803.05567 [cs].

Sui He. 2024. Prompting ChatGPT for Translation: A Comparative Analysis of Translation Brief and Persona Prompts. In *Proceedings of the 25th Annual Conference of the European Association for Machine Translation (Volume 1)*, pages 316–326, Sheffield, UK. European Association for Machine Translation (EAMT).

Wenxiang Jiao, Wenxuan Wang, Jen-tse Huang, Xing Wang, Shuming Shi, and Zhaopeng Tu. 2023. Is ChatGPT A Good Translator? Yes With GPT-4 As The Engine. *arXiv preprint*. ArXiv:2301.08745 [cs].

Arle Lommel, Hans Uszkoreit, and Aljoscha Burchardt. 2014. Multidimensional Quality Metrics (MQM): A Framework for Declaring and Describing Translation Quality Metrics. *Revista tradumàtica: traducció i tecnologies de la informació i la comunicació*, (12):455–463. Publisher: Departament de Traducció i d'Interpretació Section: Revista tradumàtica: traducció i tecnologies de la informació i la comunicació.

Jiaming Luo, Colin Cherry, and George Foster. 2024. To Diverge or Not to Diverge: A Morphosyntactic Perspective on Machine Translation vs Human Translation. *Transactions of the Association for Computational Linguistics*, 12:355–371.

Samuel Läubli, Sheila Castilho, Graham Neubig, Rico Sennrich, Qinlan Shen, and Antonio Toral. 2020. A Set of Recommendations for Assessing Human–Machine Parity in Language Translation. *Journal of Artificial Intelligence Research*, 67:653–672.

Lucía Molina and Amparo Hurtado Albir. 2002. Translation Techniques Revisited: A Dynamic and Functionalist Approach. *Meta: Journal des traducteurs*, 47(4):498–512.

Peter Newmark. 1981. *Approaches to Translation*. Pergamon Press, Oxford.

Peter Newmark. 1988. *A Textbook of Translation*. Prentice Hall, New York.

E.A. Nida. 1964. *Toward a Science of Translating with Special Reference to Principles and Procedures Involved in Bible Translating*. E.J.Brill, Leiden.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a Method for Automatic Evaluation of Machine Translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318, Philadelphia, Pennsylvania, USA. Association for Computational Linguistics.

Vikas Raunak, Arul Menezes, Matt Post, and Hany Hassan. 2023. Do GPTs Produce Less Literal Translations? In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 1041–1050, Toronto, Canada. Association for Computational Linguistics.

Ricardo Rei, Craig Stewart, Ana C Farinha, and Alon Lavie. 2020. COMET: A Neural Framework for MT Evaluation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2685–2702, Online. Association for Computational Linguistics.

Sanchez-Gijón. 2024. Towards characterizing 'generatese. In *ATISA XI Conference*.

Antonio Toral. 2019. Post-editese: an Exacerbated Translationese. In *Proceedings of Machine Translation Summit XVII: Research Track*, pages 273–281, Dublin, Ireland. European Association for Machine Translation.

Antonio Toral, Martijn Wieling, and Andy Way. 2018. Post-editing Effort of a Novel With Statistical and Neural Machine Translation. *Frontiers in Digital Humanities*, 5.

Gideon Toury. 2012. *Descriptive Translation Studies – and beyond*. John Benjamins Publishing Company. Publication Title: btl.100.

Eva Vanmassenhove, Dimitar Shterionov, and Matthew Gwilliam. 2021. Machine Translationese: Effects of Algorithmic Bias on Linguistic Complexity in Machine Translation. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 2203–2213, Online. Association for Computational Linguistics.

David Vilar, Markus Freitag, Colin Cherry, Jiaming Luo, Viresh Ratnakar, and George Foster. 2023. Prompting PaLM for Translation: Assessing Strategies and Performance. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 15406–15427, Toronto, Canada. Association for Computational Linguistics.

Jean-Paul Vinay and Jean Darbelnet. 1958. *Stylistique comparée du français et de l'anglais: méthode de traduction*. Didier, Paris.

Vered Volansky, Noa Ordam, and Shuly Wintner. 2015. On the features of translationese. *Digital Scholarship in the Humanities*, 30(1).

Longyue Wang, Chenyang Lyu, Tianbo Ji, Zhirui Zhang, Dian Yu, Shuming Shi, and Zhaopeng Tu. 2023. Document-Level Machine Translation with Large Language Models. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 16646–16661, Singapore. Association for Computational Linguistics.

Wenxuan Wang, Wenxiang Jiao, Jingyuan Huang, Ruyi Dai, Jen-tse Huang, Zhaopeng Tu, and Michael Lyu. 2024. Not All Countries Celebrate Thanksgiving: On the Cultural Dominance in Large Language Models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6349–6384, Bangkok, Thailand. Association for Computational Linguistics.

Masaru Yamada. 2019. The impact of Google Neural Machine Translation on Post-editing by student translators. *The Journal of Specialised Translation*, pages 87–106.

Yuming Zhai, Gabriel Illouz, and Anne Vilnat. 2024. Annotation Guidelines of Translation Techniques for English-French. Technical report, LIMSI, CNRS.

Wenhao Zhu, Hongyi Liu, Qingxiu Dong, Jingjing Xu, Shujian Huang, Lingpeng Kong, Jiajun Chen, and Lei Li. 2024. Multilingual Machine Translation with Large Language Models: Empirical Results and Analysis. _eprint: 2304.04675.