Università degli Studi di Modena e Reggio Emilia
DIPARTIMENTO DI STUDI LINGUISTICI E CULTURALI

Corso di Laurea Magistrale in

LINGUE, CULTURE, COMUNICAZIONE

Valutazione della qualità traduttiva nei sistemi NMT e LLMs: un'analisi comparativa di DeepL, Google Translate e ChatGPT sui testi dei Giochi Olimpici (1956–2026)

Assessing Translation Quality in NMT and LLMs: DeepL, Google Translate and ChatGPT on Olympic Games texts (1956–2026)

Prova finale di:
Elena Baldantoni

Relatrice:
Chiar.ma Prof.ssa Laura Gavioli

Correlatore:
Dott. Federico Corradini

Anno Accademico 2024-2025

SINTESI

Questa tesi esplora l'evoluzione delle tecnologie di traduzione automatica e il loro impatto sulla qualità di un testo tradotto, con particolare attenzione ai nuovi sistemi cosiddetti Neural Machine Translation (NMT) e Large Language Models (LLM). Lo studio ha l'obiettivo di valutare in che misura queste tecnologie producano traduzioni adeguate e rimodellino il ruolo del traduttore professionista. Per fare questo, la ricerca analizza le differenze tecniche tra gli output degli NMT e LLM secondo varie tipologie di errore e valuta le prestazioni di tali sistemi, utilizzando due testi, uno più recente pubblicato nel 2026 e un altro pubblicato nel 1956. In questo modo si è potuta verificare la prestazione dei sistemi di traduzione automatica, quando traducono testi che hanno diversa complessità lessicale e sintattica. Infatti, il primo testo presenta un linguaggio standard con frasi brevi, mentre nel secondo l'italiano utilizzato nella scrittura è molto più complesso, talvolta con termini desueti e frasi più lunghe.

Inizialmente si delinea una panoramica storica sulla rapida evoluzione e sul funzionamento dei sistemi di traduzione automatica, a partire dai sistemi basati su "regole", ovvero i *Rule-based machine translation systems* (RBMT) agli approcci statistici (*Statistical Machine Translation systems*), fino ai contemporanei sistemi neurali e a quelli basati sull'intelligenza artificiale. Viene esaminato inoltre il concetto di qualità applicato alla traduzione automatica, analizzando le tipologie di errore più comuni (lessicali, sintattici, semantici e contestuali) che le macchine fanno e vengono discusse anche le modalità attraverso le quali il traduttore umano interviene in questo processo, ovvero tramite le strategie di pre-editing e post-editing. La tecnica del pre-editing mira a semplificare il più possibile il testo sorgente in termini di lessico e sintassi per eliminare ogni possibile ambiguità, così da facilitare l'elaborazione del testo da parte della macchina. Tuttavia, questo processo ha uno svantaggio significativo: l'eccessiva semplificazione linguistica riduce la ricchezza stilistica del testo stesso, portando ad un sensibile impoverimento qualitativo e stilistico. Tramite il post-editing, il traduttore umano interviene sul testo dopo che la macchina lo ha tradotto, per cercare di migliorarne la qualità. Come verrà avvalorato dai risultati dell'analisi, quest'ultimo processo rimane essenziale per revisionare l'output della macchina e per correggere eventuali errori.

All'interno della tesi viene poi affrontata la questione della direzionalità della traduzione. Poiché i due testi sorgente presi in considerazione sono in italiano, che in questo caso rappresenta la L1, e la lingua di arrivo è l'inglese (L2), la direzione della traduzione sarà quella da una madre lingua ad una seconda lingua. Sebbene questa direzione sia stata spesso

considerata una pratica traduttiva sconsigliata, essa è diventata oggi sempre più comune per i traduttori, in particolar modo quando la lingua d'arrivo è l'inglese. Inoltre, vengono descritti i meccanismi di funzionamento delle tre macchine di traduzione automatica selezionate per l'analisi, ovvero i sistemi NMT DeepL e Google Translate e ChatGPT, un LLM di intelligenza artificiale generativa. Questi sistemi sono stati scelti perché sono tre degli strumenti più utilizzati oggi per produrre delle traduzioni automatiche.

Per condurre l'analisi sulla qualità degli output, lo studio combina due metriche automatiche, COMET (Crosslingual Optimized Metric for Evaluation of Translation) e TER (Translation Edit Rate) con la valutazione umana, basata sul framework MQM (Multidimensional Quality Metrics). Questo duplice approccio permette un'analisi della qualità sia quantitativa che qualitativa, in quanto le metriche automatiche forniscono dati che si riferiscono al grado di adeguatezza e fluidità delle traduzioni, mentre il framework MQM permette una classificazione più dettagliata del tipo di errori e della loro gravità.

L'analisi empirica, condotta comparando i tre output di alcune frasi-campione, rivela un netto divario prestazionale. Infatti, nel caso del testo più moderno, pre-editato e ottimizzato per le macchine, tutti i sistemi producono traduzioni generalmente accurate, anche se rimangono comunque errori lievi e incongruenze. Invece, gli output del testo del 1956 presentano una maggiore quantità di errori semantici che provengono dalla mancata comprensione del contesto da parte delle macchine. In più risultati mostrano che, mentre la traduzione di ChatGPT tende a essere più fluida e accurata, i sistemi NMT a volte producono traduzioni più letterali e apparentemente più innaturali nella lingua di arrivo. La valutazione conferma, inoltre, anche i limiti delle metriche automatiche COMET e TER, le quali, pur fornendo dati quantitativi utili, spesso non penalizzano adeguatamente gli errori di significato. Per questo motivo, il framework MQM, nonostante sia uno strumento soggettivo, è stato essenziale perché in grado di compensare i limiti di giudizio delle metriche automatiche.

Nel complesso, i risultati confermano che, nonostante i recenti e significativi sviluppi tecnologici, i sistemi di traduzione automatica non sono ancora in grado di sostituire completamente il traduttore umano. L'esperienza umana e soprattutto la modalità attraverso la quale l'essere umano è in grado di dedurre significati sottintesi e risolvere ambiguità tramite la profonda comprensione del contesto, restano indispensabili per garantire, ancora oggi, l'accuratezza di una traduzione automatica.

ABSTRACT

This thesis explores the evolution of machine translation technologies and their impact on translation quality, with particular attention to Neural Machine Translation (NMT) systems and Large Language Models (LLMs). The primary aim of the study is to assess the extent to which recent technological developments are redefining quality standards and reshaping the role of the professional translator. In addition, this research analyses the technical differences between NMT and LLM outputs in terms of types of error and evaluates the performance of the systems using two texts – one more recent, published in 2026, and another published in 1956 – with different levels of semantical and syntactical complexity.

The first chapter traces the development of machine translation, from rule-based systems (RBMT) to statistical approaches (SMT), and finally to contemporary neural and AI systems. Within this framework, particular attention is devoted to the functioning of NMT architectures, including *recurrent* and *transformer*-based models. It also examines the concept of quality in machine translation, focusing on the most common error types (lexical, syntactic, semantic and contextual) and discusses the role of human intervention through pre-editing and post-editing strategies. Pre-editing aims to simplify the source text and standardise the source language in order to facilitate machine processing, however it may also reduce stylistic richness. Post-editing, on the other hand, remains essential for correcting errors in the outputs and ensuring communicative adequacy.

The second chapter outlines the methodological framework adopted for the empirical analysis. Since the two source texts analysed are in Italian and are translated into English, the discussion begins by addressing the issue of directionality, focusing on translation from Italian (L1) into English (L2). Although translating from a mother tongue into a second or foreign language has often been considered as a problematic practice, it has become increasingly common, especially when the target language is English, a language widely required for international communication. The chapter then describes the three systems selected for the study – two NMT systems (DeepL and Google Translate) and one LLM (ChatGPT) – highlighting their different mechanisms. Particular emphasis is placed on the evaluation metrics used in the analysis: the study combines automatic metrics (COMET and TER) with human evaluation based on the MQM (Multidimensional Quality Metrics) framework. This dual approach enables both quantitative and qualitative analysis, compensating for the limitations of automatic metrics, which often fail to capture context-related errors produced by machine translation systems. The chapter concludes with a description of the methodology

employed for the empirical analysis.
The third chapter presents a comparative empirical analysis of a selected sample of sentences from the translations of two texts produced by the three systems. The analysis reveals a clear gap in their performance. In the case of the pre-edited, machine translation-friendly text all systems generally produce accurate translations, although minor errors and inconsistencies still need to be corrected. By contrast, the 1956 text poses significant challenges, as it is characterised by more complex syntax and greater contextual ambiguity, leading the systems to produce a higher number of semantic errors related to misinterpretation of the context. For this reason the post-editing effort required to correct this text is higher. The findings also show that, while ChatGPT's output tends to be more fluent and accurate, NMT systems may at times produce more literal translations that sound unnatural in the target language. The evaluation further highlights the limitations of automatic metrics COMET and TER: although they provide useful quantitative data, they often fail to adequately penalise context-related errors. For this reason, the MQM framework, despite its inherent degree of subjectivity, proves essential for a more accurate and in-depth classification of errors.
Moreover, the thesis acknowledges several limitations that may affect the generalisability of the results, including the relatively small size of the corpus analysed and the focus on a high-resource language pair (Italian – English), characterised by the availability of large digital corpora. However, these aspects also suggest directions for future research, such as investigating the quality of machine translation outputs in low-resource language pairs, i.e. languages characterised by limited available corpora for system training or exploring evaluation metrics capable of assessing texts at a broader, document-level perspective rather than focusing solely on individual sentences.
Overall, the findings confirm that, despite ongoing technological advancements, machine translation systems are not yet capable of fully replacing human translators. Human expertise, and in particular the way in which the human mind processes meaning, resolves ambiguities, and interprets implicit content, remains indispensable not only for quality assessments but also for ensuring accurate interpretation and preserving the deeper meaning of a text. Ultimately, the study supports the view that the future of translation lies in a collaborative model, in which human expertise and machine capabilities interact in a complementary rather than competitive relationship.

ZUSAMMENFASSUNG

Diese Arbeit untersucht die Entwicklung von maschinellen Übersetzungstechnologien und deren Einfluss auf die Übersetzungsqualität, mit besonderem Aufmerksamkeit auf *Neural Machine Translation* (NMT) Systeme sowie auf *Large Language Models* (LLMs). Ziel der Studie ist es, zu analysieren, inwieweit die neuen technologischen Entwicklungen die Qualitätsstandards in der Übersetzung neu definieren und die Rolle der professionellen Übersetzerinnen und Übersetzer verändern. Darüber hinaus werden die technischen Unterschiede zwischen den Outputs von NMT- und LLM-Systemen im Hinblick auf Fehlertypologien untersucht. Außerdem werden die Leistungsfähigkeiten dieser Systeme anhand von zwei Texten mit unterschiedlichem Komplexitätsgrad bewertet: einem neueren Text aus dem Jahr 2026 und einem älteren Text aus dem Jahr 1956. Dabei werden auch die Implikationen für den menschlichen Eingriff, insbesondere in Bezug auf den erforderlichen Post-Editing-Aufwand, berücksichtigt.

Das erste Kapitel zeichnet die historische Entwicklung der maschinellen Übersetzung nach – von regelbasierten Systemen (RBMT) über statistische Systemen (SMT) bis zu heutigen neuronalen und KI-basierten Systemen. In diesem Zusammenhang wird besonderes Augenmerk auf die Funktionsweise von NMT-Architekturen gelegt, einschließlich *Recurrent*- und *Transformer*-Modelle. Darüber hinaus wird das Konzept der Qualität in der maschinellen Übersetzung untersucht, wobei der Fokus auf den häufigsten Fehlertypen (lexikalische, syntaktische, semantische und kontextuelle Fehler) liegt. Zudem wird die Rolle des menschlichen Eingriffs im Übersetzungsprozess durch Pre-Editing und Post-Editing diskutiert. Während das Pre-Editing darauf abzielt, den Ausgangstext zu vereinfachen, um die maschinelle Verarbeitung zu erleichtern, kann es gleichzeitig die stilistische Vielfalt einschränken. Das Post-Editing bleibt hingegen unerlässlich, um Fehler in den Outputs zu korrigieren und die kommunikative Angemessenheit sicherzustellen.

Das zweite Kapitel beschreibt den methodologischen Rahmen der empirischen Analyse. Da die beiden untersuchten Ausgangstexte auf Italienisch verfasst sind und ins Englische übersetzt werden sollen, wird zunächst die Frage der Übersetzungsrichtung behandelt, wobei der Fokus auf der Übersetzung vom Italienischen (L1) ins Englische (L2) liegt. Obwohl die Übersetzung aus der Muttersprache in eine Fremdsprache traditionell als problematisch gilt, ist sie heute zunehmend verbreitet, insbesondere wenn die Zielsprache Englisch ist, das in vielen Kontexten stark nachgefragt wird. Anschließend werden die drei für die Studie ausgewählten Systeme vorgestellt: DeepL und Google Translate, zwei NMT-Systeme, sowie

ein LLM (ChatGPT). Ein weiterer Schwerpunkt liegt auf den verwendeten Evaluationsmethoden: die Studie kombiniert automatische Metriken (COMET und TER) mit einer menschlichen Bewertung, das MQM-Framework (Multidimensional Quality Metrics). Dieser kombinierte Ansatz ermöglicht sowohl eine quantitative als auch eine qualitative Analyse und gleicht die Einschränkungen automatischer Metriken aus, die häufig kontextbezogene Fehler nicht angemessen erfassen.

Das dritte Kapitel präsentiert die vergleichende empirische Analyse ausgewählter Beispielsätze aus den beiden Zieltexten. Die Analyse zeigt einen deutlichen Leistungsunterschied. Im Fall des neueren, *pre-edited* Textes liefern alle Systeme im Allgemeinen korrekte Übersetzungen, auch wenn kleinere Fehler und Inkonsistenzen bestehen bleiben. Der Text von 1956 hingegen stellt deutlich größere Herausforderungen dar, da er durch komplexere syntaktische Strukturen und eine höhere kontextuelle Mehrdeutigkeit gekennzeichnet ist. Dies hat dazu geführt, dass die Maschinen eine größere Anzahl semantischer Fehler produzieren, die auf ein mangelndes Verständnis des Kontexts zurückzuführen sind. Die Ergebnisse zeigen außerdem, dass die von ChatGPT erzeugten Übersetzungen flüssiger und kontextsensibler sind, während NMT-Systeme mitunter zu wörtlicher und unnatürlichen Formulierungen in der Zielsprache erzeugen. Darüber hinaus verdeutlicht die Evaluation die Grenzen automatischer Metriken wie COMET und TER, die zwar nützliche quantitative Daten liefern, jedoch häufig semantische Fehler nicht ausreichend berücksichtigen. In diesem Zusammenhang erweist sich das MQM-Framework trotz seines subjektiven Anteils als wesentlich für eine genauere Fehlerklassifikation.

Abschließend werden einige Einschränkungen der vorliegenden Arbeit aufgezeigt, die die Verallgemeinerbarkeit der Ergebnisse beeinflussen können, z.B. die geringe Größe des analysierten Korpus der Texten sowie der Fokus auf ein sogenanntes *High-Resource*-Sprachpaar (Italienisch – Englisch), für das umfangreiche digitale Korpora zur Verfügung stehen. Gleichzeitig ergeben sich daraus Ansatzpunkte für zukünftige Forschung, z.B. die Untersuchung der Qualität maschineller Übersetzungen bei *Low-Resource*-Sprachen, nämlich Sprachen mit begrenzten Trainingsdaten, sowie die Entwicklung von Evaluationsmethoden, die in der Lage sind, Texte auf Dokumentebene zu analysieren, anstatt sich ausschließlich auf einzelne Sätze zu konzentrieren.

Die Ergebnisse bestätigen, dass maschinelle Übersetzungssysteme trotz erheblicher technologischer Fortschritte noch nicht in der Lage sind, menschliche Übersetzer vollständig zu ersetzen. Die menschliche Expertise – insbesondere die Fähigkeit, implizite Bedeutungen zu erschließen, Mehrdeutigkeiten aufzulösen und kontextuelle Zusammenhänge korrekt zu interpretieren – bleibt unverzichtbar, sowohl für die Qualitätssicherung als auch für die Gewährleistung einer präzisen und sinngetreuen Übersetzung. Insgesamt unterstützt die Studie die Auffassung, dass die Zukunft der Übersetzung in einem kooperativen Modell liegt, in dem menschliche Kompetenzen und maschinelle Fähigkeiten sich gegenseitig ergänzen, anstatt miteinander zu konkurrieren.

Table of Contents

INTRODUCTION

Translation is widely recognised as a complex and interdisciplinary activity that involves not only the transfer of linguistic forms but also the interpretation of meaning within specific cultural and contextual frameworks (Holmes, 1972). Given this complexity, the use of tools designed to support the translation process has long been considered both necessary and beneficial.

In recent years, the rapid development of neural machine translation (NMT) systems and, more recently, generative artificial intelligence (GenAI) has significantly transformed the field of translation. These technologies have led to substantial improvements in the fluency and overall quality of machine-generated texts, raising important questions about the evolving role of the professional translator and the criteria by which translation quality is assessed. A number of scholars (see e.g. Kenny [2022], Bowker & Buitrago Ciro [2019], O'Brien [2022]) have reflected upon the fact that as machine translation systems become increasingly sophisticated, they are often perceived as approaching human-like performance, particularly in the production of grammatically correct and coherent outputs. However, their ability to fully capture the meaning of the words to be translated, especially when dealing with complex contents, remains a matter of ongoing debate. Although these tools are now widely accessible and commonly employed for translation between any given language pair, it remains essential to rigorously monitor the quality of their outputs, as their performance cannot be assumed to be consistently accurate.

Against this background, this study aims to investigate to what extent the evolution of NMT systems and generative AI is redefining both translation quality standards and the role of the human translator by analysing machine-generated outputs of an older Italian text published in 1956 and a more recent pre-edited one published in the current year. For this reason, some other questions that guided this research are: what are the technical differences between NMT and LLM outputs in terms of typical errors (lexical, syntactic, semantic) and how does their reliability vary between contemporary and historical texts? How much does the use of these technologies affect the actual workload of the translator in terms of post-editing effort?

In order to address the research questions and goals, the thesis is divided into three main chapters. The first chapter offers a comprehensive overview of translation technologies, exploring their evolution from rule-based machine translation to contemporary AI-driven systems. It further examines the definition of translation quality by analysing the most frequent types of errors produced by machine translation outputs and highlights the two key stages at which human intervention can enhance the process: pre-editing and post-editing. Pre-editing involves modifying the source text prior to translation to reduce ambiguity, simplify complex structures, and make implicit information explicit, thereby facilitating the processing of the text by the machine. For this reason, nowadays it is a common practice, however, pre-editing has a downside, which consists in the potential loss of distinctive linguistic, stylistic, or cultural features of a text, reducing its richness and expressiveness. Post-editing, on the other hand, entails the human revision of the translated text to correct and modify the output, ensuring coherence, fluency, and appropriateness for the target audience.

The second chapter outlines the methodological framework used to conduct the empirical analysis and addresses several key aspects. First, it discusses translation directionality, focusing on the transfer from Italian (L1) to English (L2). This direction is examined first because it is the direction chosen for the translations of this work and second, because it is a practice that has become increasingly common in contemporary translation contexts, despite having been largely overlooked in the scholarly literature. Moreover, given the pervasive influence of the English language in contemporary global communication, it is essential that automated systems perform their tasks as accurately as possible when rendering into this language. The chapter then examines the mechanisms of the neural machine translation systems employed, Google Translate and DeepL, as well as those of the large language model used, ChatGPT. Furthermore, the evaluation metrics applied for the analysis have been discussed: COMET and TER automatic metrics as well as manual assessment through the MQM framework have been chosen to evaluate the outputs. In this way, it was possible to conduct a more comprehensive and detailed analysis of the results. Finally, the chapter outlines the practical procedures implemented in the study, providing a description of the two texts used and the basis for the subsequent comparative analysis.

The third chapter describes the outputs and the results of the comparative research to identify how the engines behave when dealing with the two texts, that vary in terms of linguistic standardisation, contextual clarity, and degree of MT-friendliness. In fact, particular attention is paid to the errors produced by the systems, which, for the majority of the cases, involve contextualisation issues, as well as to the role of pre-editing in enhancing machine translation performance and the extent to which such intervention can mitigate such issues related to ambiguity and contextual underdetermination.

The findings of this study highlight a clear difference in performance between the two texts: machine translation systems tend to produce more accurate and reliable outputs when dealing with contemporary, standardised, and pre-edited texts, although errors still occur even in these cases. In contrast, they encounter greater difficulties with older texts, where linguistic variability and contextual ambiguity lead to a higher frequency of errors, including serious meaning distortions. These results provide further insight into the strengths and limitations of current machine translation technologies and contribute to the broader discussion on the continued relevance of human expertise in both the production and evaluation of translations.

Overall, this study demonstrates that, despite the significant advancements in machine translation and the increasing sophistication of automated systems, the human translator remains a central and irreplaceable figure, essential for interpreting context, resolving ambiguity, and ensuring the accuracy and appropriateness of translations.

1. MACHINE TRANSLATION

This chapter aims to outline the theoretical and technological framework within which Machine Translation (MT) has evolved, becoming one of the most ambitious and long-standing challenges of Natural Language Processing (NLP). The goal of computational linguistics has been to enable machines to understand, interpret and generate human language in a natural way and in this context, machine translation represents the fundamental testing ground for measuring the progress of Artificial Intelligence (AI), too. Within the vast field of NLP, machine translation is no longer just a theoretical challenge, but an operational tool whose effectiveness must be measured empirically in terms of accuracy, fluency and cognitive effort.

The discussion of this chapter will include a historical perspective, analysing the transition from Rule-Based Machine Translation systems (RBMT) to Statistical Machine Translation systems (SMT), up to the current dominance of Neural Machine Translation (NMT) and AI. This overview is useful for identifying the origin of the structural limitations that still characterise MT outputs today. In fact, the effectiveness of a translation system cannot be separated from a rigorous assessment of quality. As it will be analysed in the course of this work, the presence of MT errors in MT outputs requires also a change in the function of the "traditional" translator, i.e. now the translator does not disappear to yield exclusively to MT but they are evolving into figures of control and refinement of the MT outputs through the pre-editing and post-editing processes. This chapter therefore aims to provide the tools to understand how the synergy between MT systems and human sensitivity is now an essential requirement for global communication.

1.1 Natural Language Processing and Machine Translation

Human translation is understood as a complex "inter-discipline" that challenges traditional academic boundaries (Holmes, 1972). Modern scholarship in the translation field has often relied on a systematic classification that distinguished between theoretical research and practical applications. This conceptual division laid the groundwork for the field's current interdisciplinary nature since today the study of translation is no longer isolated, but rather integrated with cultural studies, digital humanities, and modern computational tools. This expansion is particularly evident in the recent and dynamic relationship between human agency and Machine Translation systems and Artificial Intelligence, which now define the cutting edge of translation research.

Machine translation (MT) consists in the automatic translation of a text from a source natural language into a target natural language, with or without human assistance (Hutchins & Somers [1992] in Bowker & Buitrago Ciro [2019]; Kenny, 2022). MT is considered to be one of the first non-numerical applications of the computers/machines after Alan Turing's discoveries on calculating machines during the Second World War (Kenny, 2022).

Technology has always been considered a valuable support for translators because the online dictionaries and the CAT tools have been able to assist them when translating, lightening their workload to some extent. However, in recent years many have witnessed huge advancements in the technological field where the unquestionable actors have become the increasingly sophisticated MT systems powered by AI and self-standing AI platforms. In addition, today these resources are free and available to anyone with a smartphone and an internet connection, whereas before technological tools were not considered essential and were accessible only to limited number of users with computers and internet access.

MT systems are one of the most prominent applications of Natural Language Processing (NLP). NLP is a branch of AI and machine learning which involves the creation of models and algorithms that enable computers to understand, interpret and generate human language, executing tasks as translation, sentiment analysis, and text summarisation (Rahaman et al., 2023). These systems attain near-human linguistic competence by processing vast amounts of language data, enabling them to perform diverse natural language tasks with remarkable accuracy. It is in this way that the most recent application of NLP, Neural Machine Translation (NMT) systems and Large Language Models (LLM) are built. In fact, the most common NMT systems e.g. Google Translate or DeepL and AI systems (ChatGPT or Gemini) can translate sentences or entire texts from one language to another in response to user inputs. These modern systems rely on the "transformer" architecture to weigh the relevance of every word in a sentence, simultaneously capturing context and semantic nuances that were previously lost in earlier systems that carried out word-for-word translations (Pérez-Ortiz et al., 2022).

Moreover, these technological systems can be combined with web browsers through the extension for the automatic translation of webpages (for instance when a user visits a website written in a language different from the one pre-set on the device, Google can suggest translating it immediately, through the Google Translate system). As Kenny (2022) further notes, MT services can be paired with automatic speech and optical character recognition or digital image processing technologies, allowing the users to translate pictures or labels in foreign languages with an app (e.g. Google Lens and its function "Translate"), as well as to translate spoken languages into written texts. Audio-visual translation represents another field in which MT is applied, e.g. the generation of subtitles in TV series, movies, videos on YouTube and captions under posts or short videos on other social media platforms[1].

Today, in a globalised world characterised by multilingual communication, machine translation has become increasingly prominent and must be considered alongside human translation practices. The advent of modern technology has changed the role of professional translators since they are no longer the only ones who have the ability to translate texts. Modern MT and AI platforms have empowered lay users to overcome linguistic barriers, facilitating the production of functional translations.

1.1.1 From the calculation machine to AI in translation

The origins of Machine Translation (MT) are deeply rooted in the early stages of computational linguistics, reflecting an extensive evolutionary process.

After the Second World War, due to the discoveries on the calculation machine made by Alan Turing, Warren Weaver introduced in his paper (Weaver's Memorandum, 1949) the idea of MT to the general public. Few years later at the Massachusetts Institute of Technology, the first conference regarding MT was held, addressing key issues as programming, funding, pre-editing and post-editing.

Moreover, during the Cold War the USA and URSS challenged each other in this field, too. They began a series of research projects aimed at translating texts or sentences between Russian and English. For the American researchers this led to the first public demonstration of how an MT system worked, the Georgetown-IBM Experiment (1954), which showed the automatic translation of a set of sentences from Russian into English. The experiment laid the foundations for further funding and for additional research in the MT field worldwide.

Nonetheless, following the initial enthusiasm, the report issued by the ALPAC (Automatic Language Processing Advisory Committee) concluded that machine translation no longer

---

[1] In 2025 Adam Moseri, Head of Instagram, announced that Instagram and Facebook would introduce the feature of Meta AI voice translation for reels on both platforms to give the possibility to creators to reach a wider audience across the globe overcoming linguistic barriers. https://www.instagram.com/p/DNiwWhoxhU0/?img_index=1

constituted a sufficiently valuable tool to warrant continued investment. As a consequence, funding and research activities in this area were largely suspended for nearly twenty years. This setback was largely due to the fact that the employed techniques were considered to be obsolete for such a highly complex process. Despite that, some research groups continued working on this field and in the 1970s some of them reached important results especially with systems working in highly-specialised areas, as the MÉTÉO system, which was able to translate weather forecasts from English into French. In Europe, the institutions adopted the SYSTRAN machine translation system which led to the development of the EUROTRA project, so as to translate sentences, and therefore entire documents, in the nine official languages of the Union (Nitzke & Hansen-Schirra, 2021).

In the 1990s the direct precursor of contemporary MT systems (especially of corpora, Google Translate, and DeepL) was developed in the form of the *Translation Memory Tool*[2], a core component of CAT Tools. Inside the TM, researchers stored previously translated documents, then the system divided the texts into smaller units (segments or sentences) and compared them with their equivalents in the source language. Therefore, if a source-language sentence was searched in the system and it matched completely or partially with stored translated sentences, the memory would propose the sentence as a possible suitable solution. Over time, as users continuously added and revised translations, the TM database expanded significantly. Thus, TM can be considered not only as the prototype of corpora, since their functioning is very similar, but also of MT systems because it laid the foundations for automated translation processes designed to operate alongside human translators.

In 1997 the first free online MT system was launched, i.e. Babel Fish on Alta Vista (Yahoo). Although these systems experienced a period of downfall at the beginning of the 21st century, it soon became clear that they were essential, as the volume of texts to be translated across multiple language combinations increased significantly due to the expansion of the internet and the globalisation[3].

A major turning point occurred in 2006 with the release of Google Translate, which was initially based on Statistical Machine Translation (SMT), and it gathered documents from the United Nations and the European Union as training material. Ten years later, the researchers transformed Google Translate SMT to a Neural Machine Translation system (NMT) so as to

[2] TRADOS (TRAnslation and Documentation Software) was the first commercial TM available, created by the researchers Jochen Hummel and Iko Knyphausen in 1984 (Nitzke & Hansen-Schirra, 2021).

[3] Even in the European Union the official languages increased significantly (from 11 to 20) and therefore the translation of the official documents couldn't rely anymore exclusively on humans.

improve its fluency and accuracy. From 2017 onwards, with the introduction of the first systems based on AI and deep learning techniques[4], MT has experienced a significant surge in performance and usability and in the same year DeepL NMT was finally launched.

Nowadays the most recent development is Artificial Intelligence (AI) a branch of computer science that designs computer programs able to solve questions and problems by acting as a human being would. It is distinguished between *narrow AI*, which is able to solve narrow problems, and *strong AI,* which is a much more complicated project, aimed at developing *general AI* or *superintelligence*, i.e. superior to human capabilities (Kenny, 2022).

The main common application of AI is Generative Artificial Intelligence (GenAI) exemplified by two major systems – Google's Gemini and OpenAI's GPT (Generative Pre-trained Transformer) systems. These models are able to understand languages and generate human-like texts in response to user prompts (Bang et al., 2023). Both the GPT system and Gemini were created by utilising advanced deep learning and Natural Language Processing techniques and are trained with Large Language Models (LLMs).[5] The GenAI systems are tools with huge potential for a wide range of application, including in the linguistic domain.

1.1.2 Types of automatic translation systems

Throughout the years different architectures of MT systems have been developed which differ from each another depending on their computational structure (Melby, 2020; Kenny, 2022). The division could be seen as follows:

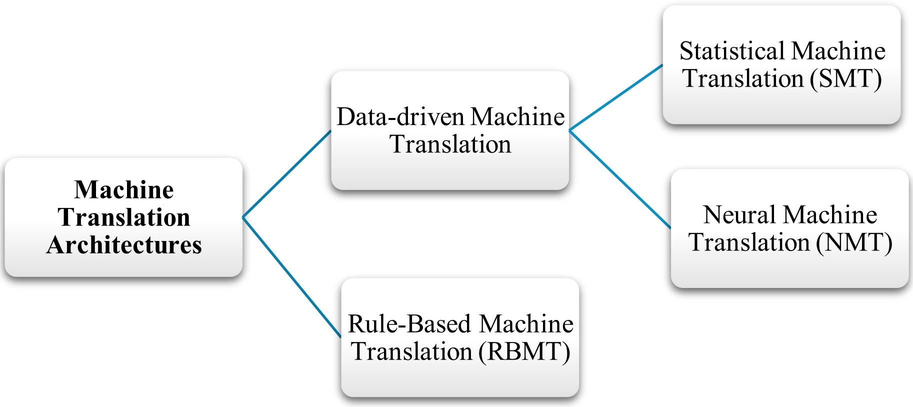

Figure 1 – Machine Translation Architectures

[4] "Deep learning is a subset of machine learning driven by multilayered neural networks whose design is inspired by the structure of the human brain. Deep learning models power most state-of-the-art artificial intelligence (AI) today, from computer vision and generative AI to self-driving cars and robotics." https://www.ibm.com/think/topics/deep-learning.

[5] Large language models (LLMs) are a category of deep learning models trained on immense amounts of data, making them capable of understanding and generating natural language and other types of content to perform a wide range of tasks. (Cole Stryker, https://www.ibm.com/think/topics/large-language-models).

The RBMT approach is based on human-made algorithms, namely rules, that permit them to produce translated sentences by using dictionaries and grammatical rules to analyse the source language and produce the translation in the target language through three possible approaches. On the contrary, SMT and NMT approaches are part of the so-called *data-driven machine translation* or *corpus-driven approaches* because “they operate by applying the results of automatically analysing training data, which consist of many source texts and their human translations” (Melby 2020:420). Therefore, the machine acquires its own knowledge by observing the problems and their possible resolutions carried out in the past.
It seems that there is no difference between RBMT and data-driven machine translation, since all computer programmes are based on algorithms, i.e. rules that need to be followed. However, the rules on which these sytems are based, are of two different types. RBMT grounds its work on numeric symbols and algebraic formulas, namely symbolic rules, predictable and controllable by human beings, whereas rules in the NMT systems are not inspectable and they are called sub-symbolic rules (Melby, 2020). As a result, NMT systems offer limited explainability compared to RBMT, making error identification and control more challenging for human users.
RBMT was dominant from the 1950s through the 1980s after which it was superseded by the SMT in the 1990s. In the 2010s the NMT emerged as the prevailing approach, and more recently AI and the GPT systems have further expanded and transformed the technological landscape.

1.1.2.1 Rule-Based Machine Translation systems

In order to perform translation, RBMT systems work with rules, as they analyse the grammatical and lexical rules of the source language, and they produce an equivalent in the target language. Bernard Vauquois was a researcher in the MT field and in 1968 he designed the “Vauquois triangle” in which he explained the three rule-based approaches used to automatically translate a text (Nitzke, Hansen-Schirra, 2021):

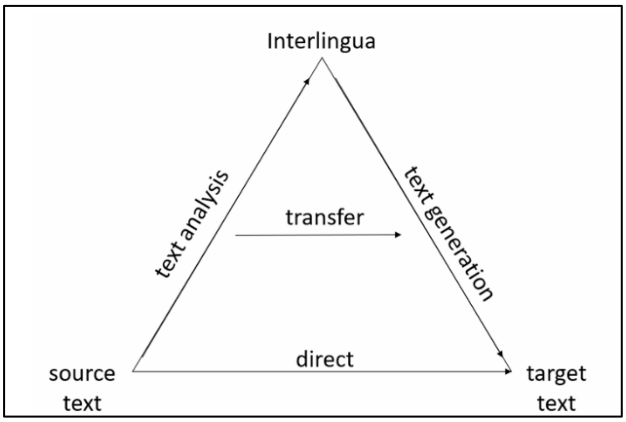

Figure 2 – The Vauquois triangle (1968, in Nitzke & Hansen-Schirra, 2021:23)

1. *Direct MT*: the direct approach consists in the direct translation from one language to another in one translation direction. This approach implies that the machine has encoded morphological rules and a dictionary of the two languages as the words of the source language are morphologically analysed and consequently replaced by the equivalent words in the target language. Initially, this process consisted exclusively of replacing words with their equivalents in the target language, and only later (in the 1950s and 1960s) researchers implemented it by introducing the functions of bilingual dictionaries and morphological and syntactic analysis of sentences.
2. *Transfer MT*: this approach is divided into three phases, namely text analysis, transfer and text generation. Firstly, the source text is analysed in depth and both a syntactic (through a syntactic tree) and sematic representation of the text are made; then the source language representation is transformed into an equivalent target language representation through a grammar that includes bilingual transfer rules. Finally, the target text is generated from the target language representation. Transfer MT offers a deeper analysis of the source text than the direct approach because it recognises the syntactic and semantic rules of the languages to reduce ambiguity, making it suitable for bidirectional and pair-specific translations. A successful example of transfer MT is SYSTRAN.
3. *Interlingua MT*: this approach tries to improve the transfer MT by removing the transfer phase. In fact, the source text is transformed into an interlingua (it represents the meaning of the source text in an abstract form) which can then be directly translated into the target language. The interlingua eliminates the need for pair-specific transfer rules because it creates a language-independent text and therefore the translation can immediately happen. Interlingua was thought to be the right system for

all translations since it provides a high degree of universality and applicability, however, this approach remained less popular than the second one.

Nevertheless, RBMT systems were expensive to develop, and it was impossible to store all the linguistic rules and general knowledge required to enable these systems to produce accurate translations. Furthermore, the systems were unable to handle lexical and structural ambiguity which are however intrinsic characteristics of natural languages, therefore they often failed to produce correct translations.

1.1.2.2 Statistical Machine Translation systems

SMT systems are grounded on training data and rely on the development of two statistical models which are "a mathematical representation of the observed data" (Kenny, 2022:36): the *translation model* and the *language model*.

The first model is a bilingual one in which *phrases,* or more precisely *n-grams* (strings of one, two, or *n* words), are analysed through a probability score. Each *n-gram* of the source language is paired with its possible translations in the target language, as identified in the training data, and they are assigned a probability score. The higher the probability, the better the translation should be. This model captures knowledge about how individual words or *n-grams* are likely to be translated into the target language.

In contrast, the *language model* shows what is likely to occur in the target language at first. It is a monolingual model of the target language based on *n-grams*. Kenny (2022:37) proposes an example of the trigram model "*I like gorgonzola*". Knowing that the *language model* tells how probable it is to find the word "*gorgonzola*" after "*I like*", she explains "It turns out to be 0.024, which means that while 'I like gorgonzola' does occur in the training data (it actually occurs four times) there are many words other than "gorgonzola" that are much more likely to follow 'I like'."

The crucial difference between SMT and RBMT is that in SMT researchers and linguists did not have to encode explicit rules as they are learned directly by the machine from the training data. Indeed, in order to make SMT systems work, the developers have engineered three phases: the *training phase*, where the machine learned the training data; the *tuning phase* where the developers set the weights to be assigned to the models to optimise translation quality; the *decoding phase* where the system produced the translated text. It functions by generating a vast array of hypothetical translations for a source-language sentence and by

identifying the one with the highest probability score. This calculation integrates the source sentence data with the system's learned patterns and assigned weights.
Nevertheless, SMT systems had its downsides. It performed best when translating texts which were very similar to the those used in the corpus as training data (for instance, if the system and its language models were trained with excerpts from cookbooks, it would have been more efficient when translating recipes). In addition, SMT systems often had the issue of word drop, where the system failed to translate one or more words within a sentence. However, until approximately 2015, SMT systems remained the predominant approach in machine translation – for example, in systems such as Google Translate and the European Commission's translation software – and played a crucial role in advancing research in both machine translation and machine learning. In fact, this line of research ultimately contributed to the development of more sophisticated architectures, most notably neural machine translation systems.

1.1.2.3 Neural Machine Translation systems

NMT systems emerged in 2015, when researchers at Stanford University developed a model that outperformed the previous SMT systems and since then NMT has been the system used for machine translation worldwide since it is considered to be the best performing approach to MT engineered so far. Initially, a widespread sense of enthusiasm pervaded the researchers, with many asserting that NMT systems were able to carry out translations which were as qualitatively good as those produced by professional human translators and that they could successfully interpret idiomatic expressions, rather than producing word-for-word translations. However, as Kenny (2022) points out, NMT systems may not translate accurately cultural-specific or idiomatic expressions since they operate by reproducing recurrent patterns learned from the training data.
The differences between SMT, RBMT and NMT lie in the models used in the more recent system. While RBMT required step-by-step instructions encoded by the developers, in NMT systems researchers present training data to the machine without specifying what to look for and the machine itself automatically searches for contextual similarities between the sentences or words (Bowker & Buitrago Ciro, 2019). Whereas SMT uses small subunits (*n-grams*), NMT systems use artificial neural networks that connect together thousands of artificial neurons with other thousands of artificial neurons. Large sets of connected artificial neurons

represent single words and their relationship with other words, allowing the system to automatically learn from vast amounts of training data through the *deep learning* techniques (Kenny, 2022). This peculiar mechanism is inspired by the functioning of the human brain.

Three layers are involved in NMT: the *input layer*, which analyses the source text, the *output layer*, which produces the target text, and other *hidden processing layers*. During the training phase, mathematical representations, namely *vectors*, and weights are assigned to the source and the target units, and this is why only the input and the output layer are known. Vectors are fixed-sized list of numbers that represent words and their relationship with other words inserted in the training data, since one can add a vector to another vector or multiply it (Kenny, 2022). The vector-based representation is also called *word embeddings*. As Pérez-Ortiz et al. (2022) explain, the embeddings are fundamental because words with similar meaning which occur in similar contexts end up having the same embeddings.

This constitutes the basis of a crucial feature of NMT systems, namely generalisation. Word embeddings project similar sentence representations and in this way the system is able to learn to translate all types of sentences, even those never processed before. Generalisation "happens when an organism which already responds to a certain stimulus in a particular way responds to similar stimuli in similar ways." (Pérez-Ortiz et al., 2022:149) and it is a skill that humans and animals have, too. In neural machine translation, generalisation means learning to translate, since neural networks can be able to produce similar outputs when fed with similar inputs, even if they are not contained in the training data.

In NMT systems two approaches have been developed: the *transformer encoder-decoder model* and the *recurrent encoder-decoder model* (Nitzke & Hansen-Schirra, 2021). In *transformer models*, the encoder converts the words of the source language in vectors to define the word embeddings, which are then processed by the decoder to generate the target text. A defining feature of this architecture is the self-attention mechanism, which enables the model to dynamically assign different weights to each element of the input sequence in relation to all other elements (Pérez-Ortiz et al., 2022). This allows the system to model complex contextual relationships and long-range dependencies without relying on sequential processing and therefore to choose the correct translation for that particular word.

The *recurrent models* rely on the encoder-decoder mechanism, too, where the encoder transforms the words into vectors and word embeddings, and the decoder pays attention to the context and produces the text in the target language. The difference is that the *recurrent model* processes words sequentially, i.e. it looks at one word at a time and passes information from previous words to the next ones. Therefore, since it processes information step-by-step, it is slower in producing the translated text and cannot handle long sentences because it can forget the meaning of previous words. Consequently, the *transformer models* are the preferred architecture for NMT systems and form the basis for the engineering of advanced AI systems such as the GPT one.

1.2 Translation quality and errors in MT outputs

As it is evident from the paragraph above, NMT and AI are the new frontiers for conducting the automatic translation process. Although they are sophisticated systems and are engineered to produce human-like products, which are fluent and accurate, they may always hide errors and imperfections. Therefore, it is important to have tools for the evaluation of MT outputs, in order to eventually provide corrections where the machine is inaccurate. Koby et al. (2014:416-417) define translation quality as such:

> "A high-quality translation is one in which the message embodied in the source text is transferred completely into the target text, including denotation, connotation, nuance, and style, and the target text is written in the target language using correct grammar and word order, to produce a culturally appropriate text that, in most cases, reads as if originally written by a native speaker of the target language for readers in the target culture."

From this definition the multifaceted nature of translation quality is highlighted, moving beyond mere linguistic accuracy to encompass the broader concept of functional and cultural similarity between source and target text. By emphasising that a high-quality translation must convey connotation, nuance, and style, the authors set a standard that transcends basic semantic transfer. From a contemporary perspective, this definition may serve as a benchmark to identify the potential limitations of NMT systems since, while these models are increasingly capable of maintaining correct grammar and word order, they may occasionally struggle with the cultural and idiomatic dimensions and the subtle nuances potentially present in a text. The evaluation process helps to analyse the correctness of all these aspects, therefore not only it assesses accuracy, fluency and style but it also measures the post-editing effort

(Lommel et al., 2014), since firstly it is crucial to detect the errors and consequently correct them so as to achieve a higher degree of translation quality.

Studies on the evaluation of MT outputs have been carried out since the 1990s, as the researchers wanted to create a standardised model for evaluating translation quality reducing the amount of subjectivity. It is even more necessary now to evaluate the MT outputs because the volume of texts and documents of all types and genres to be translated into numerous language pairs has increased significantly due to globalisation, and, as consequence, translations are progressively being handled by MT systems, too. However, users still must remain aware and understand whether a translation is actually suitable to use based on the context and the audience it is intended for.

The idea behind the evaluation process is to make a comparison between the source text, the reference text and the target text and to give a score to the translation according to its similarity to the reference text, which is considered to be a high-quality translation, usually carried out by a human translator. It is upon this process that the most used evaluation metrics are based.

The evaluation metrics help to assess how correct the MT output is, and some evaluation tools are able to detect specific errors, too. The principal categories of errors that MT systems are likely to produce are several, as outlined by Rossi and Carré (2022) and in the MQM Scorecard, which will be examined in greater detail in Chapter 2.

- Terminology: the failure to use the vocabulary that is appropriate for the subject matter and word order errors.
- Mistranslation: the source text’s meaning is misinterpreted or rendered incorrectly in the target language.
- Addition: information or concepts that are not contained in the source text have been inappropriately introduced into the output.
- Omission: specific words, phrases, or entire segments from the source text have been left out, leading to a loss of information.
- Untranslated: terms or passages that required translation are left in the source language without a valid reason.
- Grammar: violations of the grammatical, syntactical, or morphological rules of the target language.
- Punctuation: incorrect use of punctuation marks (commas, periods, quotes) that may disrupt flow or alter meaning.

- Spelling: orthographical mistakes, typos, or inconsistent spelling conventions.
- Unclear reference: occurs when a pronoun or reference (e.g. “it”, “this”, “former”) is ambiguous, making it difficult to identify the intended subject.

The errors might be grouped according to the two fundamental dimensions that govern the assessment of translation quality, namely accuracy and fluency. The former is considered to be the degree to which the target text faithfully conveys the source text’s content, and the latter concerns the stylistic and syntactic proficiency of the target text (Lommel et al., 2014). The first five error categories may refer to accuracy since they describe what changes between source and target text, while the last four may refer to the dimension of fluency because they describe the characteristics of the translated target text.

In particular errors arising from the machine’s failure to understand the context are serious issues for MT systems (as it will be discussed in Chapter 3), because such failures often stem from unresolved linguistic ambiguities. As Kenny (2022:27-29) explains ambiguity, non-isomorphism, non-compositionality and discontinuous dependencies are some linguistic phenomena that might compromise the machine’s understanding of the text and, as a result, the output produced might contain various inconsistencies. Ambiguity is the characteristic of a word or phrase to have more than one meaning (the Italian word “*pesca*” could be translated into English with the fruit “*peach*” or the activity of “*fishing*”). Non-isomorphism refers to the lack of a one-to-one correspondence between the structures of two languages (in Italian the sentence “*mi lavo le mani*” has a reflexive structure, while the sentence in English is “*I wash my hands*” with just the verb and the direct object). Discontinuous dependencies occur when two elements that are grammatically or semantically related appear separated by other words within the sentence (in Italian the question “*<u>Di</u> <u>cosa</u> parli?*” is built in such a way that the preposition and the pronoun are strictly linked and cannot be separated, while the translation in English requires that the two components disconnect, “*<u>What</u> are you talking <u>about</u>?*”). Lastly, non-compositionality occurs when the meaning of an expression cannot be deduced by adding up the meanings of its individual parts, for instance in the cases of idiomatic expressions (in Italian “*in bocca al lupo*” is an idiomatic expression that means “*good luck*” in English, and it cannot be translated word-for-word because the meaning would be completely lost).

It is true that now the NMT systems, let alone AI, are unlikely to make non-compositionality errors, especially with extremely common expressions[6], but in some cases they may still lack contextual knowledge and fail to propose a good translation. It is for this reason that humans should be aware of the potential issues and errors that MT systems may make and human intervention is necessary, especially if the text has important content information and has to be published. The post-editing and\or pre-editing processes are the two ways in which humans can interact and intervene to adjust the quality of an MT output. In the first process the text is edited after it has been translated, while in the second the text is edited before being submitted to an NMT system or AI.

1.3 Post-Editing

In response to the growing need to monitor and enhance the quality of machine translation output, the technique of post-editing (PE, or MTPE for Machine Translation Post-Editing) has emerged. This procedure consists of human intervention to systematically review, correct, and refine MT outputs, thereby allowing humans to exert quality control over the translated content. According to O'Brien (2011:197-198 in Nitzke, Hansen-Schirra, 2021:8) PE[7] "is the correction of raw machine-translated output by a human translator according to specific guidelines and quality criteria".

In this case, the translator/post-editor has to understand the intentions and the meaning of the words in the source text while simultaneously interpreting the way in which the machine has translated those sentences and correct, if necessary, the errors in the target text. The translators are able to post-edit a text because their skills enable them to draw on their world knowledge and on other translation support tools. Consequently, linguistic phenomena regarded as potential problems for MT systems, are not considered as such for humans.

Thus, PE is one of the areas in which humans and machine interact and for this reason the human intervention is a process that is able to mediate the gap of communication into two languages by editing the machine output with the least possible cognitive and temporal effort.

[6] Neither DeepL nor Google Translate have translated "In bocca al lupo" with "In the mouth of the wolf", but they used the correct English expression "Good luck".

[7] The term "Post-Editing" was already used in the past when the first MT systems were developed, it was a topic already discussed in the 1950s and 1960s. However, according to the ALPAC report, post-editing was not seen as advantageous enough compared to traditional human translation, as it did not offer significant benefits in terms of quality, time savings or reduction in the complexity of the work (Koponen, 2016).

The potential for MT to deliver high-quality output is fundamentally tied to the quality of the training data provided during the development phase. By leveraging human-made linguistic assets, such as glossaries and parallel corpora, engineers can refine NMT systems and, as the raw MT output improves, the subsequent requirement for human intervention diminishes.

However, the increasing adoption of PE has prompted requirements for a set of guidelines to achieve and ensure the expected target text quality. As Jiang et al. (2024) explain MT outputs refer to output texts generated by machine translation tools without any human edits and MT output quality is measured in terms of accuracy and complexity. TAUS[8] has developed the Quality Estimation and Automatic Post-Editing model (EPIC) and with the International Standard Organisation (ISO:18587, Translation Services – Post-editing of machine translation output – Requirements. 2017) have created sets of guidelines that concern the PE process, the output quality and the post-editor competences.

1.3.1 Light post-editing and Full post-editing

TAUS has drafted a set of guidelines to carry out PE according to two quality levels: the *good enough quality* and the *publishable quality* (Koponen, 2016:134-135). The former defines a translated text that is comprehensible and precise enough to convey the meaning of the source text without necessarily being grammatically or semantically correct. The latter describes a translated text whose quality is similar to that of a translation carried out by humans.

In addition, ISO:18587 (2017) drafted standardised requirements for PE tasks and post-editor competences. Regarding the role and the competences of the post-editor, the standard explains that they should master translation, linguistic and textual competences in both source and target languages, as well as technical competences to process the text using suitable tools, cultural competences and knowledge of the area and the theme the text deals with. Similarly to the differentiation made by TAUS, the ISO standard focuses on two types of PE, namely *light PE* and *full PE*. On the one hand, *light post-editing* refers to a rapid post-editing process where only essential corrections are made in the MT output, without the expectation of producing a text that is equivalent to a human-translation output. On the other hand, *full post-editing* is a much more thorough process, in which all the errors have to be corrected so as to achieve a result which closely resembles an output produced by humans, therefore stylistic

[8] Translation Automation User Society was established in 2004 and now it is concerned with automation and innovation in the global translation industry. https://www.taus.net/company/about-us.

quality is crucial, too. These two definitions can be aligned with the two TAUS quality levels, as *light PE* corresponds to the *good enough quality* and the *full PE* to the *publishable quality*.
According to the ISO standard the PE process should fulfil three main objectives, i.e. comprehensibility of the post-edited output, correspondence between the contents of both source and target language, the text should be following the post-editing rules provided by the translation service provider by using appropriate terminology, standard syntax, punctuation and other orthographical conventions, correct formatting and suitability for the target audience and for the purpose of the text (O'Brien, 2022).
Although these definitions provide a useful framework for understanding PE practices, they have also been subject to debate (O'Brien, 2022; Bernardini et al., 2020). Correcting only essential elements in light post-editing or trying to edit a text without any attempt in making it equal to a human-like output are challenging tasks as the boundaries between the two types of PE are blurred. In addition, comparisons between MT output and human output implicitly assume that the latter is always qualitatively superior. However, human translations do not always meet high quality standards, therefore this assumption should not be taken for granted. An additional issue lies in the fact that professional translators, post-editors or translation service providers are unlikely to state that only light post-editing has been performed, and as a result, clients and target audience may assume that a full PE has been carried out, with further implications for quality assessment and trust in professional translation services potentially leading to undervaluation or overvaluation of the translation in economic terms, too.
This lack of clear demarcation suggests that the future of post-editing will require updated technical standards as well as a shift in how quality and professional expertise are perceived and valued in an NMT-driven and AI-driven translation workflow.

1.3.2 Monolingual and bilingual post-editing

Nitzke & Hansen-Schirra (2021) clarify that it is necessary to make a further distinction within the PE process, i.e. the differentiation between monolingual and bilingual PE.
Given that today's MT systems allow anyone to translate, even non-professionals can perform a certain type of PE, namely monolingual PE. In this approach, translators or laypersons revise the MT output solely in the target language without necessarily having to access content in the source language. However, through monolingual post-editing content-related errors may not be recognised since there is no requirement to refer to the source text. By

contrast, bilingual PE process entails the juxtaposition of the target text with the source text. The post-editor must be aware not only of the semantic and syntactic accuracy in the target language, but also of the accuracy with which the meaning in the source text has been transferred into the target text, by comparing the two.

Normally content-related errors are more frequently found in monolingually post-edited texts and are considered to be the most serious ones in comparison to grammar, spelling or punctuation mistakes since they may drastically change the overall meaning of the text.

As noted by O'Brien (2022), the scholar Hans Krings (2001) observed that monolingual PE had the advantage of being much faster than the bilingual one, however, due to the high risk of undetected content-related errors, it is not considered the preferrable option. This has not changed over the years; therefore, the current recommendations advise against monolingual PE, recommending the comparison between source and target texts when undertaking a PE task (Nitzke & Hansen-Schirra, 2021).

1.3.3 The post-editor profile

The post-editor is required to develop a series of skills and competences which intersect with those of the translator, and which are essential in order to carry out post-editing on machine translation output. PE may be carried out through two types of operations, i.e. *machine translation-related processes* and *target text-related processes* (Krings 2001, in Rico & Torrejòn, 2013). The former involves reading the translated text produced by the machine to identify which elements in the sentences require intervention or reformulations. The latter is further categorised into the *target text production processes* and the *target text evaluation processes*. The first set refers to the production of a new text by maintaining the elements already contained in the text or by adding new ones to it; the target text evaluation processes involve the assessment of a negative or positive evaluation of the MT output compared to the source text.

Taking into account the processes that the post-editor has to deal with, many scholars have tried to outline the essential skills for a post-editor. Nitzke & Hansen-Schirra (2021) developed a competence model based on PACTE's translation competence model (2003), the revision competence model by Robert et al. (2017) and EMT competence framework (2022).

According to this model, the post-editor should have basic skills, as a bilingual competence so as to understand content in both the source and target language, as well as an extralinguistic competence (general knowledge) so as to comprehend the subject matter and interpret the meaning of the text accurately. A post-editor should also be able to develop a research competence, i.e. the ability to know strategically where to find missing pieces of information (online dictionaries, corpora, and others). In Rico & Torrejòn (2013) these competences are under the main category of linguistic skills and largely overlap with those required for a translator.

According to Nitzke & Hansen-Schirra (2021), the pillars of their model include error handling, MT engineering and consulting competences. Being able to recognise and correct the mistakes according to the guidelines is a crucial skill because according to them, the different MT systems generate different errors. Although the NMT system outputs are increasingly fluent and accurate, the task is still challenging because they generate errors that are difficult to distinguish. MT engineering competence requires an understanding of system features and operations, while consulting competence encompasses risk assessment and strategic decision-making during the PE process.

PE is also influenced by other surrounding factors (PACTE, 2003 and Robert et al., 2017 in Nitezke & Hansen-Schirra, 2021) such as psycho-physiological components, affinity towards technological developments, PE guidelines and the post-editor's subjectivity. It is interesting to note that Rico & Torrejòn (2013) put psycho-physiological and attitudinal competences among the core competences of a post-editor. As a matter of fact, these competences allow the post-editor to be open to lifelong learning and be prone to the new technologies, to handle the cognitive effort and stress during the decision-making process and to overcome uncertainty when post-editing (O'Brien, 2022).

Numerous academics argue that with the introduction of machine translation the boundary between translator and post-editor is not quite clear anymore. As already mentioned, it is noticeable that the basic skills of the post-editor are essentially those of a translator. Following Bernardini et al. (2020), Nitzke & Hansen-Schirra (2021) suggest that PE should not be trained separately but instead it should be assimilated into translation curricula, in late B.A. or in M.A. studies. O'Brien (2022) and Kenny (2022) support this view, with O'Brien emphasising that PE training is also valuable for linguist experts who are not trained in translation. Furthermore, Moorkens & Do Carmo (2020:42) advise that PE should be considered as integral part of the translation process "Not only because PE represents an evolution of industrial translation processes and because it fulfils the same purpose as translation (to produce a good target text in an efficient and effective way), but also because it requires advanced writing and reading skills in two different languages.”.

Although some authors (Krings, 2001 in O’Brien, 2022; Kenny, 2022) argue that post-editors must be professional translators, the ability to notice errors and edit them is not an exclusive prerogative of professional translators, but it is rather an analytical skill that can emerge where there is a strong awareness of communicative intent and technical expertise. With reference to the competences mentioned above, non-professional translators, such as translation trainees or bilingual users, may also have these skills. Possessing a certain degree of world knowledge, adequate linguistic competence in both the source and target languages, as well as familiarity with support resources such as corpora and online dictionaries, they may also be capable of performing post-editing of a text. While professional translators may be able to identify subtler errors and perhaps propose a better correction, it seems incorrect to claim that post-editing should be the exclusive prerogative of professional translators, as the necessary critical competences can also be found in non-professionals.

1.4 Controlled languages and pre-editing

A limitation of MT systems is their inability to encapsulate the full spectrum of real-world knowledge, and it is for this reason that they struggle to translate idiomatic expressions, ambiguous or long and complex sentences efficiently. Consequently, a pre-editing process has been introduced in the context of automatic translation, whereby the source text is modified prior to be processed by the machine program, in order to mitigate or prevent potential issues. Pre-editing is intrinsically correlated to post-editing as they both represent the two forms in which the human can intervene in the MT process – the more effort invested in pre-editing the less corrective work is supposed to be required during post-editing. In the words of Sànchez-Gijòn & Kenny (2022:81) pre-editing “involves rewriting parts of source texts in a way that is supposed to ensure better quality outputs when those texts are translated by machine.”.

As Bowker & Buitrago Ciro (2019) note, controlling the input before presenting it to the MT system is a concept already used in the past. An important example is the MÉTÉO system (1978), in which weather forecasts in English were translated into French on the basis of a restricted and highly specialised English sublanguage of the specific domain of weather forecasts. In fact, MT systems generally perform better when the text it has to translate belongs to a specific topic and employ a specialised lexicon (Castilho et al, 2017). In the case of MÉTÉO, a *sub-language* was used, i.e. a natural language variety characterised by a restricted lexicon that is able to explain a particular topic, such as weather forecasts terminology of the MÉTÉO system (Bowker & Buitrago Ciro, 2019).

For this reason, to optimise the pre-editing technique, the so-called *controlled languages* might be employed. They are engineered languages presenting a more restrictive lexicon, syntax and semantics in comparison with the natural language they refer to (Bowker & Buitrago Ciro, 2019; Sànchez-Gijòn, 2022). In fact, the two main characteristics of CLs are the restrictive grammar and the reduced vocabulary which serve their two main purposes, i.e. make language easier for people to understand and for machines to process. As a matter of fact, CL are divided into two categories – CL that are engineered to be more comprehensible for everyone, particularly for nonnative people, and CL that are meant for translation by an NMT system[9].

Syntactic simplification, reduction of lexical ambiguities and standardisation of terminology contribute to making a machine-friendly text, increasing the predictability of linguistic structures and facilitating computational processing. In this perspective, pre-editing is consistent with post-editing practices, as it aims to reduce both the quantity and complexity of corrections required before beginning the translation process.

The main advantages of pre-editing include greater terminological consistency, a reduction in MT output errors and an increase in overall workflow efficiency, which are particularly relevant in technical contexts characterised by high levels of standardisation. However, Controlled Languages require a significant amount of time to be engineered, and writing in CL is a very demanding skill to acquire. In addition, excessive standardisation of the source text can lead to a loss of naturalness and stylistic richness, making this practice less suitable for creative texts or texts with strong cultural content (idiomatic expressions, abbreviations, neologism, etc.). Furthermore, although pre-editing can reduce the cognitive load on the post-editor, it does not eliminate the need for human intervention but rather it integrates into a

[9] An example of the former could be the AECMA Simplified English, also known as STE (Simplified Technical English, https://www.asd-ste100.org/), a project carried out by the AECMA association (European Association of Aerospace Industries) and AIA association (Aerospace Industries Association of America) created at the end of the seventies to facilitate the maintenance of the manuals in the aeronautic field. (https://qabiria.com/en/resources/blog/controlled-language)
An example of the latter could be CTE (Caterpillar Technical English) developed by the Caterpillar manufacturer company that works with clients worldwide and had to manage 35 different languages. They invested in MT systems but rather than involving in PE processes they decided to pre-edit as much as possible the text that needed to be translated so as to improve the MT output and reduce the time and costs (Bowker & Buitrago Ciro, 2019).

continuum of strategies, depending on the quality objectives and context of use of the translation (Bowker & Buitrago Ciro, 2019; Sànchez-Gijòn & Kenny, 2022).
As reported by Sànchez-Gijòn & Kenny (2022), in 2003 O'Brien conducted a study where she tried to verify the existence of a core set of CL rules through an analysis of eight different Controlled English rule sets. She discovered that they had in common only one rule, that is, the norm that encourages the use of short sentences. Despite this lack of a core set of shared rules, her research was crucial because from the analysis she delineated the most relevant ones[10] that could be applied to Controlled English Language for producing clearer texts suitable for MT.
Following O'Brien's research, Sànchez-Gijòn & Kenny developed a (2022) series of guidelines for pre-editing aimed at improving the MT performance by enabling the production of grammatically correct translations that convey the meaning of the source text and are suited to the communicative situation of the target audience, taking into account both text function and context. The scholars divided the guidelines into three main categories: lexical choices (avoid uncommon abbreviations, lexical shifts in register and unnecessary words), structure and style (use short, simple and complete sentences that exclude ambiguity, use the active voice and an homogeneous style), referential elements (intratextual referential elements as pronouns may be mistranslated by the MT system therefore they advise to keep the sentences as simple as possible; regarding the extratextual references, any cultural aspect should be made explicit so that it becomes clear to the global audience). Moreover, Bowker & Buitrago Ciro (2019) too, created a series of recommendations so as to carry out a machine translation-friendly writing by using short sentences, avoiding wordiness, using nouns instead of personal pronouns, avoiding the use of ambiguous words, idioms or cultural references.
Furthermore, they discuss the fact that numerous service language providers and especially developers of digital products found that pre-editing a text ensures them the most effective way to communicate with the global audience since it facilitates the translation of a text into any target language, which does not imply oversimplification, but rather rendering the content understandable and transparent for a broader audience. For this reason, the suitable type of texts for pre-editing is informational, which entails unambiguous language and has the function of informing or instructing the global target audience (this will be further analysed in Chapter 2 and 3).

[10] O'Brien (2003) divided the rules into lexical, syntactic, semantic, pragmatic and text structure rules. All involve the use of a vocabulary and a sentence structure that avoid ambiguity and complexity.

Thus, pre-editing can be understood as the application of some CL or general translation friendly-writing principles, with the objective to improve MT performance while remaining embedded in a series of human-mediated translation strategies that support clear and accessible communication in multilingual and global contexts.

2. Methodological approach to NMT and AI evaluation

The evolution of contemporary textuality reflects a fundamental epistemological transition, specifically the shift from forms of writing rooted in a network of shared references and intended for a limited audience, to modes of textual production geared towards global and multilingual readership. In this framework, writing is no longer an isolated communicative act but is instead integrated into a broader process of technological mediation. The need to ensure the effectiveness of MT systems performances requires the neutralisation of potential semantic ambiguities, promoting a sort of standardisation of language (Moorkens, 2022). The syntactic simplification and denotative precision are increasingly becoming essential prerequisites for algorithmic decoding and subsequent transmission across heterogeneous linguistic and cultural contexts.

Against this theoretical backdrop, the present study is guided by the following research questions: to what extent is the evolution of NMT systems and generative AI redefining quality standards and the role of the professional translator? What are the technical differences between NMT and LLM outputs in terms of typical errors and how does their reliability vary between pre-edited contemporary and historical texts? How much does the use of these technologies affect the actual workload of the translator in terms of post-editing effort?

In order to address the research questions, this study employs a mixed-methods analytical approach, combining quantitative metrics (COMET and TER) with qualitative evaluation (MQM Scorecard). It therefore pursues a set of objectives aimed at evaluating the current state of machine translation and AI within the domain of informative texts. The primary goal is to assess the reliability of DeepL, Google Translate, and ChatGPT by comparing their linguistic outputs against a human reference. This assessment is complemented by measuring the post-editing effort to identify and categorise the most frequent translation errors through the application of the MQM (Multidimensional Quality Metrics) framework. Furthermore, the research seeks to compare the performance of NMT engines (Google Translate and DeepL) with that of Large Language Models (ChatGPT) to determine if the generative nature of AI leads to significantly different or superior qualitative results. A crucial dimension of this investigation involves analysing the impact of textual evolution, observing how the stylistic differences between a 1956 text and a contemporary, “translation-friendly” 2026 article influence the accuracy of automatic translation systems. Ultimately, these findings serve as a basis to reflect on the evolving role of the translator, discussing whether these emerging technologies act as a substitute for human translators or rather as a tool that redefines the professional figure, not only as a translator but also as a critical post-editor and supervisor.

2.1 Directionality of translation

The corpus of texts under examination, which constitutes the basis of the empirical research, is composed of two articles originally written in Italian (L1) that have been subsequently translated into English (L2) by the automatic machine translation engines. The debate on the direction of translations raises crucial questions in contemporary Translation Studies. Although traditionally regarded as problematic, L2 translation has become increasingly widespread in today's professional contexts.

Numerous scholars have reflected upon the relationship between translation quality and the concept of directionality, that is, the distinction between translations produced from an individual's mother tongue (L1) into a nonnative language (L2, L3, etc.), and those produced from a nonnative language into the individual's mother tongue. The former is commonly referred to as "inverse translation" whereas the latter as "direct translation"[11]. Translating from a non-mother tongue into the native language has always been regarded as the preferred direction, as it has been generally associated with a higher textual quality and greater reliability of the target-language output.

This belief was fostered by numerous scholars over the years and most notably through the works of Peter Newmark (1988). Indeed, he was firmly convinced that translators could only produce a natural, accurate and effective translation only when working into their first language and as a consequence, he categorically rejected translating into the L2. However, Pavlović (2008) has reported divergent results in her questionnaire survey submitted to professional translators and interpreters whose mother tongue is Croatian and who regularly translate from Croatian into English (their L2). She inquired about their profession and their views on translation directionality and discovered that it is highly common for them to translate into the second language. Moreover, she observed that translating into the second

[11] Several scholars (Pokorn [2005], Campbell [1998] and others) have rejected this taxonomy since it implies a negative judgement of translation into the L2, as Pavlović (2008) pointed out.

language is the direction preferred by the majority of them and that almost 50% receive higher rates for these translations. In addition, as she further argues, it is reasonable to assume that translations produced by native speakers into their mother tongue may be equally prone to error as translations into an L2. This assumption is confirmed by Pokorn (2005) who, through the analysis of the quality of a number of translated target texts taken as samples, found that some of them did not meet the requirements for a high-quality translation, implying that the translators demonstrated inadequacy to translate the texts in question in their L1. Nevertheless, still nowadays numerous established translation agencies and institutions hire translators who are only willing to translate into their L1[12].

The precursor to the school of thought promoted by Pokorn, Pavlović and others was Stuart Campbell (1998), the first scholar to have studied the process of translation from the mother tongue into a second language since scholars were interested only in "direct" translations. According to Pokorn (2005:35) "He argues that learning to translate is a special form of language learning and that therefore translation into a second language is not deficient *per se* but the product of developing competence". In fact, Campbell's reasoning[13] rests on two fundamental cores – the first is that translating into a nonnative language means developing skills in that language and is part of the language-learning process; his second point refers to the fact that in a multicultural world translating into a non-mother language has become a necessity (for economic reasons, international trade and high immigration rates), especially for members of linguistic minorities[14]. Thus Campbell, too, refuses the idea that translating into the nonnative language is an unnatural activity; on the contrary he is firmly convinced that it is a practice as common and widespread as translating into one's mother tongue.

According to Campbell when translators work into the L2 they select lexical items from the interlanguage, i.e. a language produced by the translators themselves. This interlanguage is derived from their understanding of the source text in the source language, yet it diverges from the established norms of the target language. The interlanguage serves as a tool that helps to analyse and comprehend how a translator works – the L2 translation is the

[12] Just as an example, this requirement is present in the Italian application page of the European Commission https://commission.europa.eu/get-involved/jobs-european-commission/working-eu-civil-service/translator-profile_it. More examples are provided in Pavlović (2008:82).

[13] Even though his research focused primarily on the educational environment, his findings and underlying arguments can be applied to the broader context of directionality in translation.

[14] Campbell (1998) studied the Arab and Vietnamese minorities in Australia and discovered that people belonging to these groups had to rely on other fellow Arab and Vietnamese immigrants, with Arab and Vietnamese mother language, in order to communicate in English. Therefore, in this case, translation occurred from the mother tongue into the nonnative language.

manifestation of their interlanguage and therefore of the way in which the translator has mentally organised the target language. This process obviously includes translation errors, which are the result of the interlanguage, too. Interestingly, the concept of interlanguage is not exclusive to human cognition but has also played a fundamental role in the development of machine translation engines. As mentioned above, early rule-based machine translation systems, for instance, used an internal interlingua (an abstract, neutral representation of the source language linking the source and target text).

Overall, the data collected by Pavlović and Pokorn together with Campbell's reflections suggest that the preference for translation into L1 is often contradicted by market dynamics and by the quality of the translated texts produced. Although many institutions continue to adhere to the traditional view that high-quality translations are those from L2 into L1, the recognition of translations into L2 as a legitimate process seems inevitable (Zanettin, 2009). In an interconnected world, where the economy is highly globalised, the traditional exclusivity of L1 translation is becoming less tenable and L2 translations are increasingly recognised as a fundamental tool for global communication, especially if the L2 under consideration is English (Stewart, 2020).

2.1.1 English as lingua franca

Any examination of directionality towards English as L2 cannot ignore the role of the language as the main *lingua franca* (ELF) in the contemporary global landscape. According to statistics from Ethnologue, in 2025 English was spoken by approximately 1.5 billion people, the vast majority of whom (1.1 billion) use it as a second language[15]. This diffusion, although rooted in historical processes of colonisation, has transformed English into a global medium of communication that functions as a "bridge" among speakers of different mother tongues. This phenomenon is by no means a recent one, as in the past other languages such as Latin, Greek or Arabic fulfilled analogous roles, i.e. operating as intermediary codes or communicative vehicles that facilitated interaction across linguistic and cultural boundaries (Munday, 2012).

[15] https://www.ethnologue.com/insights/ethnologue200/ and https://www.ethnologue.com/insights/most-spoken-language/

As Kachru (1991:180) states:

> […] English has acquired unprecedented sociological and ideological dimensions. It is now well-recognized that in linguistic history no language has touched the lives of so many people, in so many cultures and continents, in so many functional roles, and with so much prestige, as has the English language since the 1930.

In this context, Kachru's model[16] of the three concentric circles offers a fundamental key to interpreting the process of translation from Italian into English, highlighting the communicative dynamics that exist between the source language and the target language. While the *Inner Circle* identifies countries where English is the primary language, it is in the *Expanding Circle*, which includes Europe and Italy too, that the concept of ELF takes on its greatest functional relevance. Therefore, English should not be regarded as a foreign language, but rather as the main medium for the transmission of information in institutional and economic contexts (Campbell, 1998). This function is particularly evident in the process of internationalisation associated with major global events, such as the Olympic Games.

Given this global prominence, translating from Italian (L1) into English (L2) is no longer a mere linguistic exercise but a strategic operation. This challenge becomes particularly evident with the use of machine translation systems. While technology facilitates the rapid transposition of content from Italian into English, it also introduces the risk of producing outputs that are inaccurate or too closely tied to the structures of the source language. Since English is now an essential skill in international dialogue, a solid knowledge of the language is required in order to enable critical and rigorous control and evaluation over machine-translated outputs.

[16] In 1985 Kachru created a division in three concentric circles in which he explained the spread of the English language depending on the patterns of acquisition and the functional areas in which English is used (in Kachru, 1991). The *Inner Circle* comprises those countries where English is considered to be the first and primary language of the inhabitants, i.e. England, USA, Canada, Australia and New Zealand. The global spread of English originated from this core. The second circle is the *Outer Circle* which includes the countries (Pakistan, Sri Lanka, Singapore and others), that have undergone a long period of colonisation by peoples from the *Inner Circle*, during which the colonisers imposed their language, customs and culture**.** Lastly, Kachru presents the *Expanding Circle* (countries in East Asia, Mainland Europe and Latin America), which refers to those countries in which English is recognised as a universal language.

2.2 Architectural differences between NMT systems and LLMs

For the purposes of this study, it is essential to explain the mechanisms underlying the systems employed to perform automatic translations. The empirical analysis draws on two NMT systems, namely Google Translate and DeepL, and one LLM platform, namely OpenAI's ChatGPT, in order to ensure a comprehensive and representative investigation of the current technological landscape. The three engines were selected on the basis of both their research importance and their prominence in the marketplace, as they rank among the most widely employed systems for translation tasks (Sizov et al., 2024).

Although all three are used here to carry out the same task, they differ in some of their underlying architectures and operational modalities, which may in turn influence the nature and quality of the outputs they generate. NMT systems are specifically engineered to accomplish translation tasks between language pairs, while LLMs are trained on massive language corpora and are not inherently specialised translation tools. Rather translation is one of their capabilities which could be elicited through user prompting (Bang et al., 2023). By issuing specific instructions, the translator can guide the AI engine to adopt a particular persona, adhere to specific stylistic features, and maintain terminological consistency across extensive corpora. This feature is not available in NMT systems, in which the source text is directly processed and translated; in fact, the user is required only to provide the source text, the language pair and the direction of the intended translation.

Given the different volumes of training data utilised by NMT and AI systems, variations may also emerge in terms of context management since NMT systems have traditionally been geared towards sentence-level translation, while LLMs naturally support very extensive contextual frames (Sizov et al., 2024).

These structural differences are reflected in the quality of the outputs. NMTs are particularly effective for high-resource language pairs and domain-specific content, generally showing high fidelity to the source text and a lower tendency to generate content not present in the source. However, they can be too literal, less stylistically flexible and sometimes less effective in handling broad contexts or rare inputs. LLMs, on the other hand, demonstrate a greater ability to maintain consistency across multiple sentences or paragraphs, preserve tone and pragmatic register, and adapt formality and terminology according to the instructions received. Nevertheless, this flexibility carries the risk of excessive paraphrasing, over-generalisation or the addition or omission of information not explicitly contained in the source text. Indeed, as Bang et al. (2023:681) explain "LLMs are known to be susceptible to generating nonfactual, untruthful information, which is referred to as hallucination".

The comparison between NMT and LLM systems however suggests that the issue is not to establish which system is "superior" in an absolute sense. Rather, in light of these theoretical differences and characteristics, it is appropriate to examine how the three models are implemented in the main translation tools available today, to verify whether they are actually capable of delivering high-quality L2 translations and how much post-editing effort is required for the various outputs in order to eventually enhance their quality.

2.3 Performance assessment: the evaluation metrics

The adoption of evaluation metrics is an essential requirement for validating the quality of MT outputs, as MT systems have become extremely effective. A human-in-the-loop approach[17] is essential since accepting automated results without verification may undermine the quality of the final product. Thus, relying on objective assessments is a way to transform an automated process into a reliable product, since machine translation is inherently subject to unpredictable variables and potential linguistic inconsistencies.

The evaluation process could be carried out both by automatic systems (algorithms) and humans. BLEU (Bilingual Evaluation Understudy) has been one of the most used automatic evaluation systems since it was designed to resemble human judgement. The metric identifies the number of shared *n*-grams between the MT output and the reference by giving them a score from 0 to 1 – the higher the score, the closer is the MT output to the reference (Carré & Rossi, 2022). However, this metric is not complete since it focuses exclusively on lexical precision, and as consequence, other more comprehensive algorithms have been developed.

For the purpose of this work Crosslingual Optimized Metric for Evaluation of Translation (COMET) and Translation Edit Rate (TER) will be used as automatic evaluation systems and the Multi-Dimensional Quality Metrics (MQM) as the human evaluation tool. Among the high number and different typologies of automatic evaluation metrics, COMET is employed

[17] Cole Stryker defines the approach as follows: "Human-in-the-loop (HITL) refers to a system or process in which a human actively participates in the operation, supervision or decision-making of an automated system. In the context of AI, HITL means that humans are involved at some point in the AI workflow to ensure accuracy, safety, accountability or ethical decision-making". https://www.ibm.com/think/topics/human-in-the-loop

since it has been engineered on purpose to conduct evaluation analysis for NMT and AI outputs. TER measures the post-editing effort, which helps to answer the research question of how much post-editing effort is required to correct an MT output. MQM is crucial since it gives the possibility to classify the errors in an NMT and AI translation output according to their typology and severity.

The Translation Edit Rate (TER), introduced by Snover et al. (2006), measures the post-editing effort, i.e. the percentage of edits required to transform the MT output into a high-quality translation, based on a reference. Its innovation lies in its departure from the standard Levenstein distance, which is used in metrics like Word Error Rate (WER) limited to the measurement of insertions, deletions, and substitutions, by introducing a more sophisticated model of string transformation (Carré & Rossi, 2022). In fact, TER accounts for the traditional edit operations of insertion, deletion, and substitution, but its primary contribution to the field is the inclusion of shifts. By treating the relocation of an entire phrase as a single edit, Snover et al. (2006) effectively addressed the limitations of previous metrics that penalised syntactic reordering, where a displaced but correctly translated phrase would be counted as multiple errors (a series of deletions and insertions), whereas TER recognises that such errors are often the result of minor structural misalignments that require minimal cognitive effort to correct. The formula used to calculate the metric is (Snover et al., 2006:225):

$$TER = \frac{\# \, of \, edits}{average \, \# \, of \, reference \, words}$$

The metric could be measured from 0 to 1 or from 0% to 100% – the higher the score, the higher the PE effort[18].

Unbabel's COMET metric is a new model developed to measure specifically NMT outputs, therefore the translations that come from the widely used and most popular platforms, as Google Translate and DeepL, but it is also able to evaluate accurately AI outputs, too. Rei et al. (2020) explain that this model takes a step further compared to traditional metrics, since it does not exclusively take into account the lexical level of a translation in relation to its source text and reference. It evaluates translation quality by considering fluency, accuracy and the preservation of meaning. In fact, now it is crucial to take into consideration that "Modern neural approaches to MT result in much higher quality of translation that often deviates from

[18] For instance if a sentence presents a score of 70% it means that the 70% of the raw MT output must be edited to create an accurate translation, therefore the post-editor needs to be aware of the fact that the MT output was not extremely accurate.

monotonic lexical transfer between languages." (Rei et al., 2020:2685). In addition, due to the speed of evaluation, COMET is also a suitable metric for comparing different versions of machine translation systems.

An additional factor in COMET's reliability is its ability to align automatic scoring with human-led analysis. Unlike simpler metrics, it incorporates a wider range of parameters derived from human evaluation, leading to a more precise measurement of MT quality. In fact, this metrics has been engineered to maximise the correlation with human judgement since current metrics not only struggle to mirror human assessment on a segment-level but also fail to provide the sensitivity required to rank the highest-performing models. COMET calculates the similarity of vector representations of the source text, the MT output and the reference using tokens and sentence embeddings (NMT systems has the same functioning). In order to do so it leverages a neural network architecture trained on extensive datasets incorporating human judgements[19] and information from both the source text and the target-reference text, reducing the costs of manual human assessment and potentially supplementing or automating the role of human evaluators.

To complement the automated scores provided by COMET and TER, the results are further benchmarked against the detailed MQM framework. The MQM table categorises the various types of errors that may occur in a MT output, and a score is assigned based on these error categories. It is an additional mean to justify why COMET produced a specific score and offers a more comprehensive understanding of the PE effort, since "taking the PE effort into consideration is potentially even more informative because it reveals how easy or difficult it is to work with the MT output to produce a defined level of quality" (O'Brien 2022:117).

The MQM was developed as part of the European Project QTLaunchPad and is a framework for Translation Quality Evaluation (TQE) that could be applied to every kind of translation (human, NMT and AI translations). As Lommel et al. (2014) explains the MQM is not limited to a binary evaluation of the text but proposes a hierarchical and multidimensional structure that allows errors to be mapped according to their specific nature. MQM distinguishes between various dimensions such as accuracy (semantic accuracy with respect to the source text) and fluency (linguistic properties of the target text), allowing critical issues to be isolated (traditional evaluation models tend to overlap some of them). In addition, the further

[19] COMET is in compliance with the direct assessment (human evaluators assign a certain score to the output considering fluency and accuracy), the hTER (in the Human-mediated Translation Edit Rate human evaluators measure the number of editing steps required to transform the output into the post-edited version of a text) and MQM evaluation metrics.

implementation of the MQM-Core model ensures data comparability in compliance with emerging international standards[20]. This analytical approach is particularly relevant in the evaluation of MT and LLM outputs, where the ability to accurately categorise semantic distortions or register defects as opposed to purely syntactic ones, constitutes the added value. The open-access MQM scorecard[21] is a table which lists the three high-level core error dimensions, namely terminology (inconsistent use of terminology and wrong term), accuracy (mistranslation, overtranslation, undertranslation, addition, omission, untranslated) and linguistic conventions (grammar, punctuation, spelling, unclear references), and provides a space for assigning severity levels. These levels are categorised as neutral, minor, major, and critical, and as specified by this taxonomy, the errors range from the least to the most severe. The table also contains additional calculation cells that enable the user to examine the data in greater detail.

2.4 Research design

In order to evaluate the evolution and effectiveness of automatic translation engines, the study employed a combined qualitative and quantitative analytical approach applied to two journalistic and informative texts concerning the Winter Olympic Games.

The first text concerns the Cortina Winter Olympic Games held in January-February 1956 published in the Italian daily newspaper "La Stampa", and the second relates to the recently held Milano-Cortina Winter Olympic Games of February 2026[22]. The 70-year temporal distance between the two texts is particularly significant. The former was produced in a pre-digital era when the systematic preparation of materials for a global audience was not yet

[20] In particular the MQM-Core approach is based on ISO DIS 5060:2024, Translation services – Evaluation of translation output – General guidance and on ASTM WK46396: Standard Practice for Analytic Translation Quality Evaluation. https://themqm.org/

[21] The scorecard is available and free to download in the MQM website https://themqm.org/downloads/

[22] The first article was published in the evening issue of "La Stampa" on January 26, 1956, by the journalist Carlo Moriondo and was retrieved from "La Stampa" free online archive.
http://www.archiviolastampa.it/component/option,com_lastampa/task,search/mod,libera/action,viewer/Itemid,3/page,5/articleid,1586_02_1956_0022_0005_23771396/.
The second text was published on the website of Milano-Cortina Winter Olympic Games on February 6, 2026, by the editorial team of the 2026 Winter Olympic Games.
https://www.olympics.com/it/milano-cortina-2026/notizie/le-emozioni-della-cerimonia-di-apertura-di-milano-cortina-2026-i-valori-olimpici-in-una-celebrazione-diffusa .

conceived as a communicative objective, despite the international nature of the event itself. By contrast, in the latter case, addressing an international readership has become a central and explicit prerequisite in the design and dissemination of texts associated with the Olympic Games. Thus, the selection of one older text and a recently published one was motivated by the need to verify how the automatic translation systems handle different contextual and temporal references.

In both source texts the source language is Italian, and the translations have been produced into the L2, namely English. The choice of this translation direction was selected because English is the primary foreign language, and more broadly, because it constitutes the worldwide *lingua franca* and the most requested target language in the global translation market, particularly in the context of international events.

The source texts address an identical subject matter, specifically the Olympic opening ceremony. This alignment is intended to facilitate a more reliable comparison of the results obtained from the evaluation metrics, as the semantic dimension is largely equivalent, i.e. the lexical content exhibits a high degree of similarity.

The primary communicative function of the two articles is to inform the public about the opening ceremonies. The informative genre is particularly well-suited to MT systems because the texts are publicly accessible and structured as factual, chronological accounts of events, their discourse style is relatively standardised and less creative, which further enhances their suitability for MT tools. In fact, newspaper articles are typically domain-specific and are characterised by short sentences, a precise and objective lexicon, and limited use of highly creative constructions or complex idiomatic expressions. Moreover, they generally do not contain sensitive information whose mistranslation could have serious consequences (Bowker & Buitrago Ciro, 2019).

Subsequently, the two articles were translated using DeepL, Google Translate and ChatGPT. Although it would have been possible to guide ChatGPT more effectively during translation by providing additional prompt instructions (Bang et al., 2023), I chose to restrict the interaction to a simple command to translate the text into English. This decision was made to ensure comparability with NMT systems that do not support such interactive guidance.

Following a preliminary comparison of the outputs and given that an assessment of every sentence would have been methodologically inefficient (1283 words for the 1956 article and 867 words for the 2026 text), a representative sample was selected for the analysis: thirteen sentences for the former and ten for the latter. I produced a reference translation of the sentences in English, thereby yielding a reference text suitable for the application of evaluation metrics. The open-source implementations of the automatic evaluation metrics COMET and TER were then executed on the Google Colab platform, and the resulting scores were exported in tables, which will be discussed in Chapter 3.

In order to create the reference sentences, I applied the full post-editing process to the source sentences by systematically annotating potential errors or modifications, on the basis of which the MQM scorecard was subsequently completed. For the post-editing process, i.e. identification of the errors and potential improvements, online dictionaries[23], the monolingual British National Corpus (BNC), accessed via the "Sketch Engine"[24] platform, the NOW Corpus[25] and the documents on the official website of the International Olympic Committee were employed as supporting resources to obtain collocational data, thereby enhancing the lexical quality of the text.

I did not pre-edit the texts since the purpose of the research would have been compromised. It is however highly plausible that the 2026 article already underwent a process of pre-editing. The editorial strategy adopted by the organising committee was likely oriented toward the production of highly accessible content, specifically designed to be easily processed by automatic translation tools and understood by a global audience, regardless of their native language. In fact, this text appears to have been written according to translation-friendly principles. Specifically it contains short sentences, straightforward and simple syntactic and semantic structure, without subordinate clauses, ambiguous words, idiomatic expressions or culture-specific references. Moreover, it includes some anglicisms like "performers", "live" and "concept". These peculiarities can be attributed to the fact that, in this way, the information published on the website can be easily translated automatically into English. By contrast, the 1956 article has a different style. The author of the article employs longer and more complex sentences than those in the more recent text and the semantic register is markedly elevated, featuring refined and archaic lexical choices (e.g. "gremito", "vessilli", "face" which stands for "fiaccola", "formula di prammatica", "Rumenia" or "Jugoslavia").

[23] For the purpose of this study, the online dictionaries Word Reference, Merriam Webster and Collins have been used.
[24] Sketch Engine is a corpus manager and text analysis software engineered by Lexical Computing in 2003 (Stewart, 2020). Today, it offers access to over 800 text corpora, both monolingual and multilingual ones. https://www.sketchengine.eu/
[25] The NOW Corpus (News on the Web) contains 24.4 billion words of data from web-based newspapers and magazines from 2010 to the present time (the most recent update is 2026-03-15) https://www.english-corpora.org/now/

Ultimately, the objective of this study is to observe how these engines react to two very different textual registers so as to assess whether the systems maintain a high level of performance even when the semantic dimension is not “translation-friendly” but more archaic and complex, or whether their effectiveness is predominantly confined to processing contemporary, standardised content.

3. EMPIRICAL ANALYSIS

The purpose of this chapter is to present and discuss the findings of the comparative analysis conducted on the two selected texts. In order to ensure a rigorous and comprehensive evaluation, the performance of the chosen neural machine translation and AI systems – namely Google Translate, DeepL, and ChatGPT – is assessed through a hybrid methodological framework that combines quantitative metrics with qualitative error analysis.

The investigation is structured so as to highlight how these systems respond to different stylistic, linguistic and contextual challenges. More specifically, it first examines the outputs generated from the translation of the 2026 Olympic text and those related to the 1956 text, thereby accounting for variations in textual typology and historical background. Subsequently, a cross-comparison of all outputs is carried out in order to determine the extent to which factors such as textual complexity, diachronic variation, and contextual density affect the accuracy, fluency, and overall quality of the translations produced by the different systems.

The analysis is developed in two complementary stages. On the one hand, a quantitative evaluation is conducted through the application of automated metrics, namely COMET and TER, which provide a general measure of translation quality and post-editing effort. On the other hand, a qualitative assessment is performed using the MQM framework, allowing for a more fine-grained identification and classification of translation errors. This combined approach makes it possible not only to quantify performance differences, but also to better understand the underlying causes of such differences.

Particular attention is devoted to the role of contextualisation in shaping translation quality, as well as to the impact of pre-editing on the performance of the systems. By comparing pre-edited and non-edited source texts, the analysis aims to show how the explicit management of ambiguity and structural complexity can influence both the frequency and the severity of translation errors. In this sense, the chapter seeks to bridge the gap between numerical evaluation and interpretative analysis, linking empirical results to the broader theoretical issues discussed in the previous chapters.

The following sections will therefore present the results obtained through the application of the MQM framework and the automatic metrics, providing a comprehensive overview of the current capabilities and limitations of automated translation in the context of international sporting discourse.

For the sake of clarity and consistency, all automated metric values have been rounded to two decimal places. Furthermore, within the MQM Scorecard, for each error category, the distribution of errors across the translation systems has been indicated. In particular:

- *"systematic"* denotes errors occurring in the three analysed systems.
- *"shared"* denotes errors shared by two systems.

Unless otherwise specified, errors are to be understood as engine-specific.

3.1 Analysis of the MT outputs for the 2026 article

Sentence 1

Source text	**Reference text**	**DeepL**	**Google Translate**	**ChatGPT**
Sotto il segno dell'Armonia, i XXV Giochi Olimpici Invernali hanno connesso in tempo reale lo Stadio San Siro di Milano con gli altri territori parte integrante del grande evento sportivo.	Under the banner of Harmony, the XXV Winter Olympic Games connected in real-time Milan's San Siro Stadium with the other territories that are an integral part of this great sporting event.	Under the banner of Harmony, the XXV Olympic Winter Games connected Milan's San Siro Stadium in real time with other areas that were an integral part of this major sporting event.	Under the banner of Harmony, the XXV Olympic Winter Games connected Milan's San Siro Stadium in real time with the other regions integral to the great sporting event.	Under the banner of Harmony, the XXV Winter Olympic Games connected, in real time, Milan's San Siro Stadium with the other territories that were an integral part of the great sporting event.

	COMET	**TER**
DeepL	0.85	0.30
Google Translate	0.89	0.37
ChatGPT	0.83	0.26

For Sentence 1, all three translations achieved a high score in the COMET metric, while TER has given the lowest score to ChatGPT's output (meaning that less edits are required). Overall, the translations of this sentence are of a good quality. However, there is an interesting point to highlight. In this segment, a clear divergence between TER and COMET scores is observed. ChatGPT achieved the best TER (0.27), as its output closely mirrored the syntactic structure of the source text and reference translation. However, Google Translate obtained the highest COMET score (0.89), despite requiring more formal edits. This indicates that while Google Translate version deviated more from the reference's surface structure, it was recognised as highly effective at a semantic and stylistic level.

Some modifications and shifts have to be applied especially to the Google Translate output even though they do not compromise in any way the meaning of the source sentence, i.e. the wording for "*Winter Olympic Games*" that both Google and DeepL translate as "*Olympic Winter Games*" (in NOW Corpus this latter terminology occurs 4401 times while the former 7005 but in the IOC documents the most used word order is Olympic Winter Games), the positioning of the expression "*in real-time*". Moreover, the word "*territori*" has been translated by Google Translate as "*regions*" and the expression "*parte integrante di*" shortened by Google Translate with "*integral to*", which does not completely convey the intended meaning. Although "*an integral part of*" occurs more times (in the NOW Corpus it appears 111,511 times) than "*integral to*", this should not be considered as an error. It is interesting to note that both DeepL and ChatGPT used the past tense "*were*" instead of the present "*are*" before "*integral part of*", while Google Translate omits the verb. The engines failed to recognise the pragmatic context of the sentence: when the article was published the 2026 Olympics had not yet taken place, therefore the relationship between the territories and the event was a current and ongoing fact. This choice appears to be a mechanical application of tense agreement that could have been probably triggered by the preceding verb "*connected*". The systems prioritised formal grammatical consistency over the actual temporal reality of the message.

Considering the differences mentioned, the MQM Scorecard would be completed as follows:

Error Type	Neutral Errors Count	Minor Errors Count	Major Errors Count	Critical Errors Count
Terminology	0	0	0	0
Inconsistent use of terminology	0	0	0	0
Wrong term	0	0	0	0
Accuracy	0	0	0	0
Mistranslation	1	2 "*shared*"	0	0
Overtranslation	0	0	0	0
Undertranslation	0	0	0	0
Addition	0	0	0	0
Omission	0	0	0	0
Untranslated	0	0	0	0
Linguistic Conventions	0	0	0	0
Grammar	0	0	0	0
Punctuation	0	0	0	0
Spelling	0	0	0	0
Unclear Reference	0	0	0	0
Grand Total	1	2	0	0

Mistranslation is an error that occurs when the target content does not accurately or completely convey the source content. It is for this reason that three mistranslations occurred here – the former is Google Translate's "*region*" but it is signed with a neutral severity penalty, in the sense that it is still an acceptable translation (it is still rather a broad term to use in this context); the latter two are DeepL's and ChatGPT's "*were*" instead of "*are*" classified with a minor error severity, because it does not substantially compromise the comprehensibility of the content; however, when the broader context is taken into account, it may be considered as inaccurate.

Sentence 2

Source text	**Reference text**	**DeepL**	**Google Translate**	**ChatGPT**
La top model Vittoria Ceretti, che ha sfilato insieme a un gruppo di modelle che hanno indossato creazioni disegnate da Giorgio Armani, è stata scelta per rappresentare lo spirito creativo e	Top model Vittoria Ceretti, who <u>walked the runway</u> alongside a group of models wearing creations designed by Giorgio Armani, was chosen to represent the creative and contemporary spirit	Top model Vittoria Ceretti, who <u>walked the runway</u> alongside a group of models wearing creations designed by Giorgio Armani, was chosen to represent the creative and contemporary spirit	Top model Vittoria Ceretti, who <u>walked the runway</u> alongside a group of models wearing creations designed by Giorgio Armani, was chosen to represent the city's creative and contemporary	Top model Vittoria Ceretti, who <u>walked</u> alongside a group of models wearing creations designed by Giorgio Armani, was chosen to represent the city's creative and contemporary

contemporaneo della città ha portato la Bandiera Nazionale fino al palco protocollare, dove il vessillo è stato affidato al Corpo dei Corazzieri.	of the city and she carried the National Flag to the protocol stage, where it was entrusted in the hands of the Corazzieri Regiment.	of the city. She carried the National Flag to the protocol stage, where the banner was entrusted to the Corps of the Corazzieri.	spirit. She carried the National Flag to the protocol stage, where it was entrusted to the Carabinieri Corps.	spirit. She carried the National Flag to the protocol stage, where it was entrusted to the Corazzieri Corps.

	COMET	**TER**
DeepL	0.84	0.15
Google Translate	0.84	0.23
ChatGPT	0.84	0.25

In this sentence as well, the translations can be considered largely adequate according to both COMET (0.84) and TER (with DeepL performing the best translation). Apart from some minor shifts that occur in the three outputs compared to the reference text (especially in ChatGPT's output, as TER notes), there are two particular expressions to consider. Both Google Translate and DeepL have translated the verb "*ha sfilato*" with "*walked the runway*", adding the specific element of the "*runway*", probably because the engines recognised that the previous words belong to the same semantic field, i.e. fashion shows (with words as "*top model*" and "*models*") and they have interpreted "*ha sfilato*" as in a proper fashion runway. In fact, even if the general event was the Olympic Ceremony, in that moment Vittoria Ceretti and other models have actually transformed the entire stage into a runway, so as to pay tribute to the late Italian fashion designer Giorgio Armani and for this reason the reference used the same expression. On the contrary ChatGPT has translated the verb with "*walked*", using a broader term.

Furthermore, it is interesting to note how the three systems translated "*corpo dei Corazzieri*", which in Italian is a specialised unit belonging to Carabinieri. Google Translate is the system which has deviated the most from the source text and reference translation since it has translated the name as "*Carabinieri corps*", however not obtaining the lowest scores for TER metrics. While the translation is not completely incorrect, it is nonetheless imprecise, given that the Corazzieri Regiment is institutionally a part of the Carabinieri Corps. In the source text, and specifically at that moment of the opening ceremony, the focus was exclusively on the Corazzieri, as they were the ones designated to raise the Italian flag. By substituting the

name of the specific regiment with the name of the general corps, the engine loses the pragmatic precision intended in the original description of the opening ceremony.
DeepL and ChatGPT left the original Italian name of "*Corazzieri*" as in the source text and attached it to the word "*corps*". This is not wrong but to be extremely precise the Corazzieri are a cavalry regiment rather than a corps, which is bigger and composed by multiple units. For this reason, the correct translation would be "*Corazzieri Regiment*".

In accordance with COMET and TER scores, the MQM Scorecard would be completed as follows:

Error Type	Neutral Errors Count	Minor Errors Count	Major Errors Count	Critical Errors Count
Terminology	0	3 "systematic"	0	0
Inconsistent use of terminology	0	0	0	0
Wrong term	0	0	0	0
Accuracy	0	0	0	0
Mistranslation	0	0	1	0
Overtranslation	0	0	0	0
Undertranslation	1	0	0	0
Addition	0	0	0	0
Omission	0	0	0	0
Untranslated	0	0	0	0
Linguistic Conventions	0	0	0	0
Grammar	0	0	0	0
Punctuation	0	0	0	0
Spelling	0	0	0	0
Unclear Reference	0	0	0	0
Grand Total	1	3	1	0

ChatGPT's translation of "*ha sfilato*" with "*walked*" has been classified as an undertranslation with neutral severity, since the word is semantically less specific than the corresponding term used in the source Italian text. Google Translate's rendering of "Corazzieri Corps" as "*Carabinieri Corps*" has been classified as a major mistranslation because the word used in the MT output does not faithfully reflect the source term, thereby compromising the accuracy of the description of the Ceremony.
Finally, Google Translate and ChatGPT's use of "*Corps*" or DeepL's "*the Corps of*" instead of "*Regiment*" have been annotated as inaccurate terminology errors with a minor error severity level, since they do not compromise the understandability of the text, but diminish the technical precision and overall terminological quality of the output.
Another noteworthy observation is that the original Italian sentence is relatively long, and the systems determined that it was appropriate to segment it into two separate sentences by inserting a full stop in the middle of the extended Italian sentence. This could be considered as an addition or a punctuation error; however, it should not be interpreted as a mistake or a negative element. On the contrary, it may enhance the comprehensibility of the phrase.

Sentence 3

Source text	**Reference text**	**DeepL**	**Google Translate**	**ChatGPT**
Le 92 delegazioni, partendo dalla Grecia e chiudendo con l'Italia, hanno sfilato nei siti più vicini alle proprie sedi di gara, trasformando la geografia italiana in un unico palcoscenico condiviso.	The 92 delegations, starting with Greece and ending with Italy, paraded at the sites closest to their competition venues, transforming the Italian geography into a unified shared stage.	The 92 delegations, starting with Greece and ending with Italy, paraded in the locations closest to their competition venues, transforming the Italian landscape into a single shared stage.	The 92 delegations, starting in Greece and ending in Italy, marched at the sites closest to their respective competition venues, transforming the Italian geography into a single shared stage.	The 92 delegations, starting with Greece and closing with Italy, paraded at the venues closest to their competition sites, turning Italy's geography into a single shared stage.

	COMET	**TER**
DeepL	0.83	0.14
Google Translate	0.83	0.25
ChatGPT	0.79	0.32

The analysis of this segment highlights one of the most significant paradoxes in MT and its evaluation through neural and statistical metrics. The core of the issue lies in the translation of the phrase "*partendo dalla Grecia e chiudendo con l'Italia*", which refers to the specific parade order of the Olympic delegations.

In this context, Google Translate's output ("*starting in Greece and ending in Italy*") represents a substantial semantic and pragmatic error. The use of the preposition "*in*" suggests a physical geographical journey between the two nations, completely ignoring the ceremonial logic in which the delegations parade sequentially inside the stadium. Despite this evident contextual misinterpretation, the quantitative data provides a distorted view of the actual quality: Google Translate achieved a COMET score of 0.83 and a TER of 0.25, appearing to be similar to DeepL's output and at the same time outperform ChatGPT.

Google Translate's COMET score (0.83) highlights a significant oversight by neural metrics toward contextualisation and world knowledge. COMET evaluates quality based on the proximity of semantic vectors (Rei et al., 2020). Because key terms such as "*delegations*," "*Greece*," "*Italy*," "*starting*", "*sites*", "*competition venues*" are present, the system considers the translated sentence plausible. However, it fails to perceive that a minor prepositional shift (*in* vs. *with*) utterly compromises the logical coherence of the event.

Conversely, ChatGPT provided a logically fluent translation ("*starting with Greece and closing with Italy*"), yet it was paradoxically penalised by both metrics, recording the lowest COMET (0.79) and the highest TER (0.32). TER penalised ChatGPT for its formal distance from the reference translation, e.g. "*closing*" instead of "*ending*" or the syntactic restructuring of "*Italy's geography*" instead of "*the Italian geography*", or using "*competition sites*" instead of the most used form in the Olympic documents "*competition venues*". This demonstrates how TER functions as an indicator of lexical similarity.

Ultimately, this sentence is emblematic as it proves that automated metrics may reward a logically flawed translation if it maintains strong surface-level similarity to the source or reference text. This confirms that human evaluation remains a highly reliable tool for identifying pragmatic and errors in context that bypass statistical and algorithmic calculation. In fact, for this sentence the MQM Scorecard does penalise Google Translate's performance by marking the parenthetical element "*starting in Greece and ending in Italy*" as a mistranslation with a major error severity, since in this case it makes the sense of the sentence distorted. ChatGPT's c*ompetition sites*" instead of "*competition venues*" has been classified as a wrong term but with neutral error severity, because "*venues*" is the correct word when dealing with sporting events and competitions; however, the term used by ChatGPT does not alter the general meaning of the sentence.

Error Type	Neutral Errors Count	Minor Errors Count	Major Errors Count	Critical Errors Count
Terminology	0	0	0	0
Inconsistent use of terminology	0	0	0	0
Wrong term	1	0	0	0
Accuracy	0	0	0	0
Mistranslation	0	0	1	0
Overtranslation	0	0	0	0
Undertranslation	0	0	0	0
Addition	0	0	0	0
Omission	0	0	0	0
Untranslated	0	0	0	0
Linguistic Conventions	0	0	0	0
Grammar	0	0	0	0
Punctuation	0	0	0	0
Spelling	0	0	0	0
Unclear Reference	0	0	0	0
Grand Total	1	0	1	0

Sentence 4

Source text	**Reference text**	**DeepL**	**Google Translate**	**ChatGPT**
Scelta simbolica dell'Italia che, per rappresentare l'unità del Paese, ha schierato eccezionalmente quattro portabandiera: due a Milano. Arianna Fontana e Federico Pellegrino e due a Cortina, Federica Brignone e Amos Mosaner.	Italy made a symbolic choice by fielding four flagbearers to represent national unity: two in Milan, Arianna Fontana and Federico Pellegrino, and two in Cortina, Federica Brignone and Amos Mosaner.	This was a symbolic choice by Italy, which, to represent the unity of the country, exceptionally fielded four flag bearers: two in Milan. Arianna Fontana and Federico Pellegrino and two in Cortina, Federica Brignone and Amos Mosaner.	This was a symbolic choice by Italy, which, to represent the country's unity, exceptionally fielded four flag bearers: two in Milan, Arianna Fontana and Federico Pellegrino, and two in Cortina, Federica Brignone and Amos Mosaner.	As a symbolic choice, Italy—representing national unity—exceptionally fielded four flag bearers: two in Milan, Arianna Fontana and Federico Pellegrino, and two in Cortina, Federica Brignone and Amos Mosaner.

	COMET	**TER**
DeepL	0.69	0.53
Google Translate	0.70	0.40
ChatGPT	0.73	0.37

The analysis of this segment reveals a relatively high level of performance across all three systems, as evidenced by both COMET and TER scores, although certain differences emerge when combining quantitative and qualitative evaluation. ChatGPT achieves the highest COMET score (0.73) and the lowest TER score (0.37), indicating the closest alignment with the reference translation (and, consequently, with the source text) in terms of both adequacy and fluency. Its output is indeed stylistically refined and structurally coherent, employing punctuation (e.g. dashes) to enhance readability while maintaining the original meaning. For instance, it accurately renders the peculiar initial expression of the source sentence "*Scelta simbolica dell'Italia…*" which structure is often used in Italian journalistic writing.

Google Translate and DeepL also produced acceptable translations (COMET 0.70 / TER 0.40 and COMET 0.69 / TER 0.53, respectively), albeit with some deviations from the reference.

Nonetheless, both systems adhered very closely to the source sentence, resulting in overly literal renditions that compromised idiomatic fluency in English. In fact, for instance, the expression "*This was a symbolic choice by Italy*" is markedly unnatural in English. Moreover, they differ from the reference translation because the former employed the expression "*the unity of the country*" and the latter used the Saxon's genitive "*the country's unity*". In addition, DeepL has incorrectly segmented the sentence, as demonstrated by the incorrect insertion of a full stop after "*Milan*", which disrupts the syntactic continuity of the phrase. This is not an error generated autonomously by the system: the same punctuation error is already present in the source text, and the engine, having produced a highly literal translation, has simply replicated it. It is nevertheless classified an error because neither ChatGPT nor Google Translate have made the same mistake, correctly identifying that a full stop at that position would be semantically and syntactically inappropriate.

Overall, in the human evaluation conducted using the MQM scorecard, only a single minor punctuation error was identified in the output generated by DeepL. Nevertheless, it is important to emphasise that, despite the relatively low number of explicit errors in this sentence, the translations produced by Google Translate and DeepL cannot be regarded as fully adequate, as they are excessively literal. This tendency toward literal translation represents a critical limitation of NMT systems.

Sentence 5

Source text	**Reference text**	**DeepL**	**Google Translate**	**ChatGPT**
All'interno dello stadio San Siro, la Torcia "Essential" è stata protagonista di una staffetta simbolica: prima portata da tre tedofore, tra cui Paola Egonu, poi affidata a due successivi gruppi di tre atleti italiani per l'uscita dallo stadio	Inside San Siro Stadium, the Torch, <u>named "Essential",</u> took centre stage in a symbolic relay: initially carried by three torchbearers, including Paola Egonu, then passed to two subsequent groups of three Italian athletes for <u>its exit</u> from the	Inside the San Siro stadium, the <u>"Essential" Torch</u> was the star of a symbolic relay: first carried by three torchbearers, including Paola Egonu, then entrusted to two successive groups of three Italian athletes <u>to leave</u> the	Inside the San Siro Stadium, the <u>"Essential" Torch</u> took center stage in a symbolic relay: first carried by three torchbearers, including Paola Egonu, then entrusted to two successive groups of three Italian athletes <u>for the exit</u>	Inside San Siro Stadium, the <u>"Essential" Torch</u> took center stage in a symbolic relay: first carried by three torchbearers, including Paola Egonu, then passed to two subsequent groups of three Italian athletes for <u>their exit</u> from the

accompagnati dalla performance di Andrea Bocelli.	stadium, accompanied by a performance by Andrea Bocelli.	stadium accompanied by a performance by Andrea Bocelli.	from the stadium accompanied by a performance by Andrea Bocelli.	stadium, accompanied by a performance by Andrea Bocelli.

	COMET	**TER**
DeepL	0.68	0.33
Google Translate	0.72	0.17
ChatGPT	0.78	0.09

In this segment, ChatGPT demonstrates a superior ability to handle the translation. With an impressively low TER of 0.09, its output requires minimal post-editing, aligning with the source and reference sentences. The high COMET score (0.78) further validates ChatGPT's performance.

Conversely, DeepL recorded the lowest performance in both metrics. This can be attributed to a slip in register, specifically for the use of the term "*the star of*" to translate "*protagonista*", which feels perhaps overly informal for an Olympic ceremony. This example highlights how LLMs (like ChatGPT) can sometimes outperform traditional NMT engines (like DeepL) in capturing the register required.

A slight variation between the reference and all three outputs concerns the placement of the name "*Essential*". While the reference translation treats it as a distinct title through an appositive structure ("*the Torch, named 'Essential'*"), all three engines automatically applied the standard English attributive rule, placing the name before the noun ("*the 'Essential' Torch*"). Although the engines' outputs are grammatically correct, they slightly alter the focus: the reference emphasises the naming of the object, so as to make it clear to the readers, whereas the machines treat "*Essential*" as a descriptive attribute, to maintain a certain degree of synthesis (a key feature of these systems). This structural shift, however, had a minimal impact on the TER scores, as the metric penalises local word reordering less severely than lexical omissions, substitutions or undertranslations.

With respect to the MQM scorecard two types of errors were identified. The first is an instance of undertranslation, classified with neutral error severity, for failing to clarify that "*Essential*" is the name of the Torch, but this does not impede the audience's understanding of the sentence. This error was produced by all three engines. The second is a minor unclear reference, related to ChatGPT's use of the plural possessive "*their exit*" instead of the singular "*its exit*". In context, ChatGPT appears to refer simultaneously to both the athletes and the Torch's exit; nonetheless, given that the grammatical subject of the sentence is the Torch, the singular possessive pronoun is perhaps more appropriate in order to foreground the Torch's movement. However, ChatGPT's choice is understandable, as it is clear that both the torch and the athletes have left the stadium. DeepL and Google Translate remained more faithful to the source text by using "*to leave*" and "*for the exit*".

Error Type	Neutral Errors Count	Minor Errors Count	Major Errors Count	Critical Errors Count
Terminology	0	0	0	0
Inconsistent use of terminology	0	0	0	0
Wrong term	0	0	0	0
Accuracy	0	0	0	0
Mistranslation	0	0	0	0
Overtranslation	0	0	0	0
Undertranslation	**3 *"systematic"***	0	0	0
Addition	0	0	0	0
Omission	0	0	0	0
Untranslated	0	0	0	0
Linguistic Conventions	0	0	0	0
Grammar	0	0	0	0
Punctuation	0	0	0	0
Spelling	0	0	0	0
Unclear Reference	0	**1**	0	0
Grand Total	3	0	0	0

Sentence 6

Source text	**Reference text**	**DeepL**	**Google Translate**	**ChatGPT**
Così la Fiamma Olimpica ha unito la dimensione urbana di Milano con la maestosità delle Dolomiti patrimonio UNESCO.	In this way, the Olympic Flame united Milan's urban dimension with the majesty of the Dolomites, a UNESCO World Heritage Site.	Thus, the Olympic Flame united the urban dimension of Milan with the majesty of the Dolomites, a UNESCO World Heritage Site.	Thus, the Olympic Flame united the urban dimension of Milan with the majesty of the Dolomites, a UNESCO World Heritage Site.	In this way, the Olympic Flame united Milan's urban dimension with the majesty of the UNESCO-listed Dolomites.

	COMET	**TER**
DeepL	0.94	0.29
Google Translate	0.94	0.29
ChatGPT	0.88	0.29

In this segment, DeepL and Google Translate outperform ChatGPT in terms of COMET (0.94 vs 0.88), despite all three systems share an identical TER (0.29). This discrepancy highlights a specific qualitative compromise: while ChatGPT achieved better syntactic fluency by using the Saxon genitive ("*Milan's urban dimension*", as it was chosen in the reference), it was penalised by the neural metric for its less formal rendering of "*patrimonio UNESCO*" as "*UNESCO-listed*". This observation is confirmed by data from the NOW Corpus, which indicate that the expression "*UNESCO-listed*" occurs 2,435 times, whereas the phrase "*UNESCO World Heritage Site*" appears 20,734 times. For this reason, in the MQM Scorecard this would be classified as a neutral or minor error (specifically inconsistent use of terminology).

In contrast, DeepL and Google Translate maintained the official institutional designation ("*a UNESCO World Heritage Site*"), which aligned more closely with the register of the reference translation, too. This case demonstrates that while LLMs like ChatGPT often prioritise synthesis, they may occasionally sacrifice terminological precision – a factor that COMET, being sensitive to semantic nuances, is able to detect more accurately than a purely distance-based metric like TER.

Overall, among the ten evaluated sentences, this one obtained one of the highest scores, supporting the observation that the systems exhibit superior performance on short sentences with relatively simple syntactic structure (O'Brien, 2003; Bowker & Buitrago Ciro, 2019; Sànchez-Gijòn & Kenny, 2022).

Two further points are worth noting, as they involve errors produced by almost all three machine translation systems under consideration. The first concerns Google Translate and DeepL rendering of "*braciere olimpico*" as "*Olympic brazier*", which can be classified as a terminology error. Although the term "*brazier*" is semantically plausible, it does not correspond to the conventional and officially established expression "*Olympic cauldron*", resulting in a less accurate and domain-appropriate translation and demonstrating the tendency of NMT systems to produce literal renderings. The second issue relates to the translation of "*Milano-Cortina 2026*" as "*Milan-Cortina 2026*", which constitutes an error in the handling of named entities. In this case, all three systems fail to preserve the official designation of the event, where the Italian form "*Milano*" is maintained even in the English versions of the Olympic documents.

While neither of these errors significantly compromises overall comprehension, they reduce the terminological precision and institutional accuracy of the output, highlighting the systems' limitations in adhering to domain-specific conventions and official branding protocols.

3.2 Analysis of the MT outputs for the 1956 article

The second part of this analysis focuses on the 1956 text, a corpus that presents different metrics compared to the contemporary 2026 data. The lower COMET scores and the fluctuations in TER values observed in this section are not necessarily indicative of a failure in translation quality but rather highlight the inherent limitations of automated metrics when processing non-contemporary source. Since these algorithms are primarily trained on modern linguistic datasets, they may struggle to validate the formal register, archaic syntax, and technical terminology. Consequently, the MQM scorecard becomes an even more critical tool in this chapter, as it allows for a qualitative distinction between necessary stylistic modernisation and actual translation inaccuracies that automated systems fail to interpret correctly.

Sentence 7

Source text	**Reference text**	**DeepL**	**Google Translate**	**ChatGPT**
I Giochi Olimpici si sono iniziati ufficialmente questa mattina alle 10.10, quando il Presidente della Repubblica nello Stadio del Ghiaccio, popolato di concorrenti e gremito di folla, si è alzato per pronunciare la formula di prammatica: «Dichiaro aperti i VII Giochi Olimpici Invernali di Cortina d'Ampezzo celebranti la XVI Olimpiade dell'era moderna».	The Winter Olympic Games officially began this morning at 10:10, when the President of the Republic <u>stood up in the Ice Stadium, filled with athletes and spectators,</u> to pronounce the <u>customary oath</u>: "I declare open the VII Winter Olympic Games of Cortina d'Ampezzo, celebrating the XVI Olympiad of the modern era".	The Olympic Games officially began this morning at 10.10 a.m., when the President of the Republic <u>stood up in the Ice Stadium, filled with competitors and packed with crowds,</u> to pronounce the <u>customary formula</u>: "I declare open the VII Winter Olympic Games in Cortina d'Ampezzo, celebrating the XVI Olympiad of the modern era".	The Olympic Games officially began this morning at 10:10, when the President of the Republic <u>stood in the packed Ice Stadium</u> to pronounce the <u>formal oath</u>: "I declare open the VII Olympic Winter Games of Cortina d'Ampezzo, celebrating the XVI Olympiad of the modern era."	The Olympic Games officially began this morning at 10:10, when the President of the Republic, <u>in the Ice Stadium filled with competitors and packed with spectators, rose</u> to pronounce the <u>customary formula</u>: "I declare open the VII Winter Olympic Games of Cortina d'Ampezzo, celebrating the XVI Olympiad of the modern era."

	COMET	**TER**
DeepL	0.73	0.18
Google Translate	0.70	0.25
ChatGPT	0.72	0.27

For this sentence COMET scores are relatively high, demonstrating that the MT outputs are of generally good quality, as they preserve the overall meaning of the source sentence. The lowest COMET score is obtained by Google Translate's output, which has been presumably penalised for its choice to compress the information contained in the first part of the sentence. In order to emphasise the fact that the Cortina Ice Stadium was very crowded, the author of the article employs the following description: "S*tadio del Ghiaccio, popolato di concorrenti e gremito di folla*". Whereas DeepL and ChatGPT translated this part by maintaining the same syntactic structure of the source, Google Translate condenses it to "*the packed Ice Stadium*". This behaviour highlights a tendency in certain MT systems to summarise the information as much as possible whenever possible, rather than reproducing it in a more explicit and fully elaborated form. The systems demonstrated adequate performance in rendering the archaic expression "*formula di prammatica*": DeepL and ChatGPT translated it with "*customary formula*", Google Translate rendered it with "*formal oath*". In this specific context of the Olympic Games, the translation "*oath*" is more appropriate, as the term "*formula*" does not appear in the official IOC documents, while "*oath*" is frequently found.

Regarding the TER scores, DeepL receives the highest score, because it is more syntactically and semantically similar to the reference. By contrast, Google Translate and ChatGPT obtain the lowest scores because the former lacks of the information mentioned above and the latter introduces structural modifications to the sentence (for instance "*in the Ice Stadium filled with competitors and packed with spectators*" is rendered as a parenthetical element rather than a locative complement, as in the source text, reference and in the other two MT outputs).

Consequently, in the MQM Scorecard a minor omission from Google Translate and minor mistranslations for DeepL's and ChatGPT's use of "*formula*".

Error Type	Neutral Errors Count	Minor Errors Count	Major Errors Count	Critical Errors Count
Terminology	0	0	0	0
Inconsistent use of terminology	0	0	0	0
Wrong term	0	0	0	0
Accuracy	0	0	0	0
Mistranslation	0	**2 *"shared"***	0	0
Overtranslation	0	0	0	0
Undertranslation	0	0	0	0
Addition	0	0	0	0
Omission	0	**1**	0	0
Untranslated	0	0	0	0
Linguistic Conventions	0	0	0	0
Grammar	0	0	0	0
Punctuation	0	0	0	0
Spelling	0	0	0	0
Unclear Reference	0	0	0	0
Grand Total	0	3	0	0

Sentence 8

Source text	**Reference text**	**DeepL**	**Google Translate**	**ChatGPT**
Poi la bandiera dai cerchi di cinque colori è salita lentamente sul pennone principale e l'artiglieria ha sparato tre salve di cannone che a lungo si sono ripercosse tra le gole dei monti incombenti.	Then the flag with the five-coloured rings slowly rose on the main flagpole and the artillery fired three cannon salvoes that echoed for a long time through the gorges of the surrounding mountains.	Then the flag with the five coloured circles slowly rose on the main flagpole and the artillery fired three cannon salutes that echoed for a long time in the gorges of the looming mountains.	Then the five-colored circled flag slowly rose to the main mast and the artillery fired three cannon salutes that echoed for a long time through the gorges of the looming mountains.	Then the flag with the five-colored rings slowly rose on the main flagpole, and the artillery fired three cannon salvos that echoed for a long time among the surrounding mountain gorges.

	COMET	**TER**
DeepL	0.71	0.37
Google Translate	0.65	0.53
ChatGPT	0.75	0.20

The analysis of this segment reveals a more marked divergence in performance across the three systems, particularly when comparing COMET and TER scores. ChatGPT achieves the highest COMET score (0.82), indicating a superior ability to preserve the overall adequacy and fluency of the translation. Its output is indeed highly natural and stylistically coherent, with only minor lexical variation from the reference (e.g. "*salvos*" instead of "*salvoes*" and "*among*" instead of "*through*"), which does not significantly affect meaning. However, its TER score (0.27) is higher than DeepL's, suggesting a greater number of surface-level variations from the reference despite its higher overall quality.

DeepL, on the other hand, records a lower COMET score (0.68) but the best TER score (0.18), indicating a closer formal correspondence to the reference translation. Its output remains generally accurate, although some lexical choices, such as "*circles*" instead of "*rings*" and "*salutes*" instead of "*salvoes*," slightly reduce terminological precision and stylistic appropriateness.

Google Translate shows the weakest performance in this sentence, with the lowest COMET score (0.61) and a relatively high TER score (0.27). The translation contains a more evident structural issue in the phrase "*five-colored circled flag*," which appears syntactically unnatural and deviates from standard English usage, thus negatively affecting both fluency and clarity.

In addition, Google Translate and DeepL have translated "*tre salve di cannone*" with "*three cannon salutes*", while ChatGPT and the reference used "*three cannon salvo(e)s*". On Sketch Engine "*salute(s)*" is intended as a verb that conveys a greeting, while "*salvo(e)s*" is paired with "*cannon*" or in general more linked with the artillery semantic field. However, if research is conducted on Google, the most used terminology to explain blank shots used in ceremonial contexts such as the Olympic Opening ceremony is "*cannon salutes*". For this reason, it will not be considered as an error.

The expression "*gole dei monti incombenti*" has been translated faithfully by ChatGPT ("*in the surrounding mountain gorges*") and by DeepL and Google Translate as "*looming mountains*". Although on Sketch Engine the word "*mountain*" is often paired with "*surrounding*" and not with "*looming*", the translation is not considered as an error. However, the adjective "*looming*" conveys a sense of imposing or even potentially threatening presence, which may not fully align with the celebratory tone of the source text.

Overall, this example highlights a divergence between the evaluation metrics: while TER has rewarded closer surface similarity to the reference, COMET appears more sensitive to semantic adequacy and overall fluency. As a result, ChatGPT emerges as the most effective system in qualitative terms, despite not achieving the lowest TER score, confirming that higher textual naturalness does not always correspond to minimal edit distance from the reference.

In the MQM scorecard, the annotated errors would correspond to Google Translate's minor mistranslations of "*five coloured circled flag*" and "*mast*," as the latter is a term predominantly used in a naval context.

Error Type	Neutral Errors Count	Minor Errors Count	Major Errors Count	Critical Errors Count
Terminology	0	0	0	0
Inconsistent use of terminology	0	0	0	0
Wrong term	0	0	0	0
Accuracy	0	0	0	0
Mistranslation	0	2	0	0
Overtranslation	0	0	0	0
Undertranslation	0	0	0	0
Addition	0	0	0	0
Omission	0	0	0	0
Untranslated	0	0	0	0
Linguistic Conventions	0	0	0	0
Grammar	0	0	0	0
Punctuation	0	0	0	0
Spelling	0	0	0	0
Unclear Reference	0	0	0	0
Grand Total	0	2	0	0

Sentence 9

Source text	**Reference text**	**DeepL**	**Google Translate**	**ChatGPT**
Questi ultimi sembrano troppi; facciamo notare che sono invece pochi rispetto alla reale quantità degli «ufficiali» scesi a Cortina al seguito dei concorrenti.	<u>The latter seem too many;</u> we note, however, that they are actually few compared to the real number of team officials who <u>have come</u> to Cortina <u>to follow the athletes</u>.	<u>The latter seem too many;</u> we would point out that they are actually few compared to the real number of “officials” who <u>came</u> to Cortina <u>accompanying the competitors</u>.	<u>The latter number may seem excessive;</u> it should be noted that they are actually few compared to the actual number of "officials" who <u>descended on</u> Cortina <u>to accompany the competitors</u>.	<u>The latter seem too many;</u> we note, however, that they are actually few compared with the real number of “officials” who <u>have come</u> to Cortina <u>in the athletes’ wake</u>.

	COMET	**TER**
DeepL	0.71	0.37
Google Translate	0.65	0.53
ChatGPT	0.75	0.20

In this example, both COMET and TER produced a consistent ranking, with the ChatGPT translation outperforming DeepL and Google Translate's outputs. This convergence is particularly informative, as it highlights a case in which semantic adequacy and surface-level similarity align. The ChatGPT version closely mirrors the reference both structurally and lexically (e.g., "*we note, however, that they are actually few*"), resulting in a low number of required edits and thus a favourable TER score. At the same time, it employs natural and idiomatic expressions such as "*in the athletes' wake*", which enhance fluency and semantic precision, leading to a higher COMET score. By contrast, DeepL translation, while semantically accurate, introduces less idiomatic phrasing (e.g., "*we would point out*", "*accompanying the competitors*"), increasing the edit distance from the reference. Google Translate's output exhibits more substantial deviations, as "*may seem excessive*" and lexical choices as "*descended on Cortina*", which affect both semantic fidelity and structural alignment.

In the MQM framework "*descended on Cortina*" could be considered as a minor mistranslation since it does not hinder the understandability of the sentence, however it has a different nuance than the one present in the source text. The verb "*to descend on*" typically conveys the idea of arriving in large numbers, often suddenly and with a somewhat negative or intrusive connotation[26]. By contrast, the Italian source phrase "*al seguito dei concorrenti*" refers to officials accompanying the athletes in a neutral sense.

Sentence 10

Source text	**Reference text**	**DeepL**	**Google Translate**	**ChatGPT**
Verso il termine tre squadroni: i russi, gli americani, gli italiani, che sono i gruppi più numerosi. All'apparire dei nostri l'urlo della folla diviene	<u>Toward</u> the end we see <u>the three largest delegations</u>: the Russians, the Americans, and the Italians. When our team appeared, <u>the crowd erupted in a thunderous roar</u>; we	<u>Towards</u> the end, <u>there are three large teams</u>: the Russians, the Americans and the Italians, <u>which are the largest groups</u>. When our team appeared, <u>the</u>	<u>Towards</u> the end, <u>three teams emerge</u>: the Russians, the Americans, and the Italians, <u>who are the largest groups</u>. When our team appears, <u>the roar of the crowd escalates</u>	<u>Toward</u> the end <u>come three large squads</u>: the Russians, the Americans, and the Italians, <u>the largest groups of all</u>. At the appearance of our

[26] https://www.wordreference.com/enit/descend

tempesta: non sappiamo se gli amplificatori degli apparecchi televisivi hanno potuto registrarlo con esattezza. Guida i nostri il saltatore Nilo Zandanel.	do not know whether television amplifiers were able to record it accurately. Leading our athletes is the ski jumper Nilo Zandanel.	crowd's roar became a storm: we do not know if the television amplifiers were able to record it accurately. Our team was led by the jumper Nilo Zandanel.	into a storm: we don't know if the television amplifiers were able to accurately record it. Leading our team is high jumper Nilo Zandanel.	own team, the roar of the crowd becomes a tempest; we do not know whether television amplifiers were able to record it accurately. Leading our athletes is the ski jumper Nilo Zandanel.

	COMET	TER
DeepL	0.79	0.53
Google Translate	0.78	0.59
ChatGPT	0.80	0.47

In this example, the variation between "*toward*" and "*towards*" reflects a difference between American and British usage rather than a semantic or lexical error. While this distinction has no impact on meaning and is therefore ignored by semantic metrics such as COMET, it may slightly affect TER due to surface-level mismatch. In addition, at the beginning of the sentence, the reference merges two semantically related segments of the source ("*tre squadroni*" and "*i gruppi più numerosi*") into a single clause ("*we see the three largest delegations*") in order to avoid repetition and enhance fluency. This operation is not observed in the systems' outputs, which adhere more closely to the source structure and consequently retain redundant elements.

Both COMET and TER indicate that ChatGPT's translation outperforms DeepL and Google Translate outputs, although all three systems receive relatively low TER scores, suggesting notable word divergence from the reference. ChatGPT's version achieves the highest COMET score (0.80) and the lowest TER (0.47), reflecting a better balance between semantic adequacy and structural similarity to the reference and source text. In particular, it closely matches the reference in segments such as "*Leading our athletes is the ski jumper Nilo Zandanel*", while also preserving a faithful and appropriate rendering of "*l'urlo della folla diviene tempesta*" as "*the roar of the crowd becomes a tempest*". This expression was quite hard to render: in the reference it is translated as "*the crowd erupted in a thunderous roar*", which maintains the same effect of a sudden and very loud sound, through the terms "*erupted*" and "*thunderous roar*". Perhaps the engines' renditions with words like "*tempest*" or "*storm*" paired with the verb "*become*" are weaker.

The use of ChatGPT's "*squads*" represents an inadequate lexical choice in this context, as it fails to capture the institutional meaning conveyed by "*delegations*" in the reference. While not entirely incorrect, it introduces a slight semantic mismatch and a more informal tone, which may negatively affect both COMET and TER scores. In addition, the official documents of the International Olympic Committee (IOC) explicitly employ the terms "*delegations*" and "*teams*."

DeepL shows deviations both lexically and structurally, as in "*there are three large teams*" and "*our team was led by the jumper*", "*towards*" and "*the crowds roar became a storm*" which increase the distance from the reference.

Google Translate performs slightly worse in COMET (0.78) and in TER (0.59) due to multiple issues, including syntactic shifts and a critical lexical error for "*high jumper*", which misinterprets "*saltatore*". In this context the author refers to the athlete Nilo Zandanel who competes in ski jumping rather than in athletics. DeepL too almost makes the same mistake by translating the word with just "*jumper*", which is not precise but could be accepted. The two NMT systems were deviated by the term "*saltatore*" of the source text which constitutes an ambiguous element for the engines. Unable to recognise the broader context, they assumed that the best translation would be "*jumper*" or "*high jumper*", which is the literal translation of the word. By contrast, ChatGPT was successfully able to infer by the context of the sentence that the athlete performed a winter sport and thus produced the correct translation.

For these reasons, the MQM Scorecard includes two minor mistranslations for ChatGPT's "*squads*" and for DeepL's "*jumper*", and one critical wrong term for Google Translate's "*high jumper*".

Error Type	Neutral Errors Count	Minor Errors Count	Major Errors Count	Critical Errors Count
Terminology	0	0	0	0
Inconsistent use of terminology	0	0	0	0
Wrong term	0	0	0	1
Accuracy	0	0	0	0
Mistranslation	0	2	0	0
Overtranslation	0	0	0	0
Undertranslation	0	0	0	0
Addition	0	0	0	0
Omission	0	0	0	0
Untranslated	0	0	0	0
Linguistic Conventions	0	0	0	0
Grammar	0	0	0	0
Punctuation	0	0	0	0
Spelling	0	0	0	0
Unclear Reference	0	0	0	0
Grand Total	0	2	0	1

Sentence 11

Source text	**Reference text**	**DeepL**	**Google Translate**	**ChatGPT**
Ora una figurina lo segue sul palco: la nostra Giuliana Minuzzo, medaglia di bronzo ad Oslo, campionessa di discesa.	Now another athlete follows him on stage: our very own Giuliana Minuzzo, bronze medallist in Oslo, downhill champion.	Now a figure follows him onto the stage: our Giuliana Minuzzo, bronze medallist in Oslo, downhill champion.	Now a figurine follows him onto the stage: our own Giuliana Minuzzo, bronze medalist in Oslo and downhill champion.	Now a small figure follows him onto the platform: our Giuliana Minuzzo, bronze medalist at Oslo, downhill champion.

	COMET	**TER**
DeepL	0.85	0.33
Google Translate	0.83	0.44
ChatGPT	0.79	0.56

In this example, the COMET and TER scores reveal a divergence between semantic adequacy and surface-level similarity. DeepL achieves the highest COMET score (0.85) and the lowest TER (0.33), indicating that its translation is both semantically more acceptable and closer to the reference. Although the use of "*figure*" does not accurately convey the meaning of "*figurina*" in this context, it remains a semantically plausible, human-related term and does not introduce a strong contextual inconsistency. The reference sentence uses the terms “*another athlete*” because the author is talking about a second athlete who is about to enter the stage. By using the word “*figurina”*, the author wants to convey that the athlete appeared as a tiny figure in the distance, emphasising her small stature or\and how far away she was from the author who was describing the scene. Perhaps this nuance is lost in the reference, however, for the purposes of the translation, it might have been better to make it clear that the person walking onto the stage was, first and foremost, an athlete.

By contrast, Google Translate’s “*figurine*” results in a clear semantic error, as it denotes an inanimate ornamental object, thus significantly diverging from the intended meaning. This explains its relatively high TER (0.44). The ChatGPT version, despite being fluent, receives a lower COMET score (0.79) and the highest TER (0.56). This is due to the addition of “*small*”, which introduces unsupported information, as well as to structural differences such as “*onto the platform*” instead of “*on stage*”, or “*our Giuliana Minuzzo*” instead of “*our very own Giuliana Minuzzo*” of the reference. These variations increase the edit distance from the reference and are penalised by TER.

Moreover, unlike the term “*jumper*” in the preceding example, the expression “*downhill*” in this context functions as the technical designation for the alpine skiing discipline. Consequently, the addition of a qualifying prefix such as “*alpine skiing*” is semantically redundant. While contemporary usage might occasionally require clarification to distinguish it from “*mountain bike downhill*”, such potential ambiguity is absent within the Olympic and broader context of the 1950s. Therefore, when Giuliana Minuzzo competed in major sporting events, “*downhill*” unambiguously referred to the alpine skiing event, thereby rendering any further specification unnecessary for the audience.

Furthermore, the use of “*platform*” instead of “*stage*” by ChatGPT is not regarded as an error and similarly, the form “*medalist*,” produced by both ChatGPT and Google Translate, is not a typographical mistake but rather reflects the standard American orthography.

For these reasons, the MQM scorecard identifies one critical wrong term (in compliance with COMET score), i.e. the rendering of “*figurina*” as “*figurine*”, produced by Google Translate.

Error Type	Neutral Errors Count	Minor Errors Count	Major Errors Count	Critical Errors Count
Terminology	0	0	0	0
Inconsistent use of terminology	0	0	0	0
Wrong term	0	0	0	1
Accuracy	0	0	0	0
Mistranslation	0	0	0	0
Overtranslation	0	0	0	0
Undertranslation	0	0	0	0
Addition	0	0	0	0
Omission	0	0	0	0
Untranslated	0	0	0	0
Linguistic Conventions	0	0	0	0
Grammar	0	0	0	0
Punctuation	0	0	0	0
Spelling	0	0	0	0
Unclear Reference	0	0	0	0
Grand Total	0	0	0	1

Sentence 12

Source text	**Reference text**	**DeepL**	**Google Translate**	**ChatGPT**
Forse è più emozionata che alla partenza di una gara di grande impegno. Afferra con la destra un lembo della bandiera italiana e mentre gli altri vessilli si piegano sino a sfiorare la pista, pronuncia la formula del rito:	Perhaps she is more emotional than at the start of a highly demanding race. With her right hand she grasps a fold of the Italian flag, and while the other flags bow until they nearly touch the track, she pronounces the oath:	Perhaps she is more excited than at the start of a major competition. She grabs a corner of the Italian flag with her right hand and, while the other flags are lowered to touch the track, she recites the ritual formula:	Perhaps she's more excited than she is at the start of a challenging race. He grasps a corner of the Italian flag in his right hand, and as the other flags fold until they touch the track, he pronounces the ritual formula:	Perhaps she is more emotional than at the start of a highly demanding race. With her right hand she grasps a fold of the Italian flag, and while the other banners bow until they nearly touch the ice, she pronounces the ritual formula:

	COMET	**TER**
DeepL	0.77	0.38
Google Translate	0.75	0.50
ChatGPT	0.79	0.10

The results of the automated metrics for this sentence yielded consistent outcomes since Google Translate has committed three serious errors in its output, and in fact it has been penalised by both COMET (0.75, which is the lowest) and TER (0.50, which is the highest value, thereby indicating that its output requires substantial modification). By contrast, ChatGPT has produced the best translation output.

In this sentence, the word "*emozionata*" must be rendered into the target language. DeepL and Google Translate use the term "*excited*" while ChatGPT "*emotional*". In this context, the term "*emotional*" was preferred over "*excited*" to better convey the solemnity of the moment, since "*excited*" typically denotes high energy and anticipation and "emotional" captures the profound sense of being moved or touched by the ritual. Thus, here the metrics seems to have correctly evaluated the outputs.

Nevertheless, leaving aside the fact that DeepL has translated the verb "*afferra*" with "*grabs*" which is weaker than "*grasps*" that on the contrary denotes a firm, intentional grip, it is Google Translate that made a substantial error that could hinder the audience understanding of the text. Google Translate misgendered the subject by using masculine pronouns instead of the required feminine forms. Despite the context clearly referring to a female athlete, the translation algorithm opted for masculine pronouns (*he/him*) instead of feminine ones (*she/her*). Google Translate's preference for masculine pronouns reflects what Moorkens (2022:136), through Vanmassenhove's (2019) research, describes as an "algorithmic bias that exacerbates existing gender bias in training data […] with genders assigned inconsistently, even within a single segment". Although the context requires a feminine subject, the system defaults to the more "statistically common" masculine form.

Lastly, the three engines rendered the expression "*la formula del rito*" as "*the ritual formula*". However, in the official Olympic documentation, the term "*formula*" in this particular context does not occur, whereas the appropriate expression in this case is "*Olympic oath*". In relation to the expression "*the ritual formula*" DeepL selected the verb "*recites*" instead of "*pronounces*" which appears both in the reference and in Google Translate and ChatGPT's outputs.

In fact, in the MQM Scorecard Google Translate's error is categorised as critical, while the others have all neutral or minor severity levels.

Error Type	Neutral Errors Count	Minor Errors Count	Major Errors Count	Critical Errors Count
Terminology	0	0	0	0
Inconsistent use of terminology	0	0	0	0
Wrong term	0	***3 "systematic"***	0	0
Accuracy	0	0	0	0
Mistranslation	**1**	***2 "shared"***	0	**3**
Overtranslation	0	0	0	0
Undertranslation	0	0	0	0
Addition	0	0	0	0
Omission	0	0	0	0
Untranslated	0	0	0	0
Linguistic Conventions	0	0	0	0
Grammar	0	0	0	0
Punctuation	0	0	0	0
Spelling	0	0	0	0
Unclear Reference	0	0	0	0
Grand Total	1	5	0	3

The most serious issues are the three critical mistranslations related to gender pronouns. By using masculine pronouns (*he/his*) for a female subject who had just been described with the feminine pronouns, the engine creates deep ambiguity. This is especially problematic as a male athlete had been mentioned shortly before, leading the reader to a complete referential confusion regarding the identity of the protagonist of the action.

From a technical perspective, the choice of "*ritual formula*" instead of the official "*oath*" is categorised as a minor wrong term, as it fails to adhere to the formal Olympic taxonomy found in institutional documents. Furthermore, DeepL and Google Translate translation of "*emozionata*" as "*excited*" represents a minor mistranslation, as it leans toward a generic state of excitement rather than the deeper emotional state implied by the Italian source in a ceremonial context. Lastly, the use of "*grabs*" instead of the more precise "*grasps*" is marked as a neutral mistranslation; while it conveys the general action, it lacks the nuanced control and solemnity required for handling a national flag during an official rite.

In this instance COMET failed to significantly penalise Google Translate for its severe errors, suggesting that even advanced neural metrics can overlook critical shifts in meaning if the overall fluency remains high.

3.3 Discussion

Overall, the engines performed better on the 2026 text, exhibiting fewer and less serious errors. Consequently, the post-editing effort would be evidently lower than what would be required for the 1956 article.

Considering the twelve sentences examined from both the 2026 and the 1956 articles, a total of fifteen errors were identified in the former, and nineteen errors in the latter. DeepL produced nine errors, Google Translate sixteen, and ChatGPT nine. Of the errors identified in Google Translate, five were classified as critical and two as major, whereas the remaining errors across DeepL and ChatGPT were limited to neutral or minor severity levels. In fact, these errors would require only minor corrections, as they can be considered as slight modifications[27]. The more serious errors are instead contextual ones, which could instead require more post-editing effort because they might compromise the text's comprehensibility for the intended audience.

Neural machine translation systems are often described as being broadly inspired by the architecture of the human brain, as they rely on artificial neural networks that resemble, in a simplified form, the functioning of human neurons. This resemblance has led to the assumption that such systems might eventually approximate human-like reasoning, including at the linguistic level. Indeed, contemporary automated translation technologies, including Large Language Models, demonstrate remarkable performance and are capable of producing fluent and contextually plausible outputs in many cases. However, despite these advancements (and acknowledging that the development of such technologies is still ongoing) it remains unlikely that they will fully replicate the complexity of the human cognitive processes.

Human language processing involves the continuous integration of vast amounts of information from both linguistic and extralinguistic sources. This process is inherently complex and depends on the dynamic interaction of multiple factors, among which context plays a central role (Mason, 2006). From this perspective, context can be seen as the fundamental element that shapes meaning, enabling interlocutors or readers to interpret

[27] For instance in the first sentence of the 1956 text the expression "*formula di prammatica*" has been translated as "*customary formula"* and "*formal oath*", with the latter being correct but slightly inaccurate; in the 2026 article "*braciere olimpico*" has been translated by DeepL and Google Translate as "*Olympic brazier*" a solution that is somewhat imprecise with respect to the official Olympic taxonomy ("*Olympic cauldron*" would have been the correct choice).

nuances, resolve ambiguities, and infer implicit content that is not explicitly encoded in the linguistic form.

The centrality of context can be further explained through Carston's underdeterminacy thesis (in Mason, 2006), according to which the linguistic meaning encoded in an utterance is insufficient on its own, to fully determine its interpretation. Meaning derives from the interaction between linguistic input and contextual inference. In written communication, as considered in the present analysis, the context of a given sentence emerges from its relationship with the surrounding textual environment, namely the sentences that precede and follow it within the same text. Meaning is thus constructed through a network of formal and semantic relations that link single phrases into a coherent whole. This has significant implications for machine translation: while neural systems and large language models can process large amounts of textual data, they lack the inferential mechanisms required to bridge the gap between encoded meaning and intended meaning, which is precisely where many translation errors originate. A major factor influencing a machine's capacity to accurately interpret contextual information is linguistic ambiguity, which can, in some cases, pose a substantial challenge for the processing engines.

Machine translation systems operate primarily on the basis of statistical regularities and learned representations of linguistic patterns and for this reason, despite their advanced capacities, they still encounter significant difficulties in processing such contextual dependencies (Pérez-Ortiz et al., 2022). They can take into account broader textual segments, but they do not truly "understand" context in the way humans do, nor can they reconstruct the inferential processes underlying human communication. Moreover, as Vu et al. (2024) explain, when context is excessively long, translation engines often fail to pick the right information. Instead of considering the whole text and the broader context, they prefer to choose simpler statistical patterns that make predictions[28] easier even though these may generate reduced accuracy.

In the sentences analysed above, the three engines have produced several errors, particularly in the outputs of the 1956 article, as they had to process a set of peculiar lexical items that do not belong to the standard language. Consequently, in this case predictions were not straightforward to make, requiring the systems to solve more ambiguities that may have arisen during the processing of the Italian text. Several issues remained largely unresolved because

[28] According to Vu et al. (2024), predictions refer to the lexical and grammatical choices made by the engines during the translation process. They often choose immediate and frequent statistical patterns, ignoring the more complex information present in the document's context.

the engines lack the ability to reason across the full context, particularly Google Translate. Compared to DeepL and ChatGPT, this system was the one who produced more major or critical errors, all concerning its ability to process the context.

In the 1956 article, Google Translate rendered the term in *Sentence 10* "*saltatore*" as "*high jumper*", once again failing to account for the broader discursive context. The source text did not provide explicit disambiguation, as the author of the article employed a rather generic term to indicate Nilo Zandanel's sporting discipline; however, it was inferable from the surrounding context that he was not an athletics jumper. In *Sentence 11*, Giuliana Minuzzo is referred to as "*figurina*" in the source text, meaning that she was much smaller than her teammate who went up on stage before her, or it could also refer to a figure appearing in the distance, which the author therefore cannot see clearly at first glance. Google Translate was not able to understand this nuances in the meaning of the Italian word and therefore it automatically translated the term literally. However, "*figurine*" in English does not have the same denotative meaning, rather it refers to an actual static object. Furthermore, Google Translate rendered the personal pronouns in *Sentence 12*, which should have been referred to the female athlete Giuliana Minuzzo, with "*he/his*" because shortly before another athlete was mentioned and he was a male (Nilo Zandanel). Thus, the system may have confused the reference of these pronouns, unable to learn it from the context. There is another contextual error in this sentence, this time committed by both Google Translate and DeepL – they have translated "*emozionata*" with "*excited*" instead of "*emotional*", which does not convey the right meaning.

For the 2026 article, contextualisation errors were less frequent, of comparatively lower severity, and were also found in the outputs produced by DeepL and ChatGPT. In *Sentence 1* these systems translated the verb "*be*" with the past tense ("*…areas/territories that were an integral part of…*") incorrectly implying that the event was already over at the moment of speaking. Two further contextualisation errors have been made by Google Translate in *Sentence 2* and *Sentence 3* – the former is the translation of "*Corpo dei Corazzieri*" as "*Carabinieri Corps*", in which the engine produced an overgeneralisation of the term, loosing accuracy[29]. The latter is the expression of *"…partendo dalla Grecia e chiudendo con l'Italia…"* at the moment of the nation parade, which the system rendered as *"…starting in Greece and ending in Italy…"*. It seems that Google Translate implies that the parade physically started in Greece and physically ended in Italy, which is not the right assumption.

[29] Shortly after the source text cites "Carabinieri", therefore Google Translate might have taken this translation as a valid prediction for both terms.

From the context (and the previously translated sentences) it should have understood that the author was describing the order in which the nations taking part in the Games would be entering during the opening ceremony in Milan, Cortina and the other Italian competition venues.

Therefore, what can be inferred from the analysis is that the Large Language Model ChatGPT may handle contextual information somewhat more effectively than traditional Neural Machine Translation systems, allowing for more coherent and nuanced outputs. Moreover, it did not exhibit any instances of so-called "hallucination", a phenomenon recognised as a potential behaviour in large language models (Bang et al., 2023). Nevertheless, they still fall short of fully replicating human-like inferential and pragmatic reasoning, and errors arising from underdetermined context remain common. Moreover, DeepL and Google Translate produced more literal outputs in comparison with ChatGPT's ones[30].

This persistence of contextual failure suggests that the optimisation of machine translation cannot rely solely on algorithmic improvements but sometimes may require a modification of the input itself. In light of these limitations, the role of pre-editing emerges as a crucial factor in enhancing the performance of machine translation systems. The pre-editing process is helpful to modify a source text prior to automatic translation in order to reduce ambiguity, simplify syntactic structures, and make implicit information more explicit. In doing so, it effectively compensates for some of the shortcomings related to contextual underdeterminacy, by reshaping the input into a form that is more easily processable by machine translation systems.

The comparative analysis conducted in this study clearly supports this view: when comparing the machine translation output of a pre-edited text with a non-edited one, a significant difference can be observed both in the number and in the severity of errors[31]. This suggests that, while machine translation systems struggle to perform the inferential processes required to resolve underdetermined meanings, they can nonetheless benefit from inputs in which such inferential work has been, at least partially, anticipated and made explicit by the human author.

From this perspective, pre-editing can be interpreted as a form of guided mediation between human communicative complexity and machine processing limitations. Rather than

[30] This could be seen in *Sentence 4* and for other expressions present in the other phrases of the articles. For instance, to translate "*braciere*", they used "*brazier*" instead of "*cauldron*", which is the official name when dealing with the Olympic context or Google Translate's literal rendering of "*figurina*" with "*figurine*", and so on.

[31] Google Translate's critical errors have all been found in the 1956 article.

overcoming the intrinsic limitations of automated systems, it adapts the source text to align more closely with their operational mechanisms (Bowker & Buitrago Ciro, 2019). As a result, the improvement in translation quality is driven not just by how the machines are now able to "understand" language, but also by the extent to which they are provided with texts of lower structural and semantic complexity, which in turn demand less sophisticated internal processing and interpretation. This further confirms that the challenges posed by context and underdeterminacy are not fully cancelled but strategically managed through human intervention.

Another issue that emerges from the analysis concerns the automatic evaluation metrics, such as COMET and TER. Although they provide useful quantitative indications of translation quality, they often exhibit limitations in detecting and appropriately penalising errors that arise from contextual misinterpretation. Such errors can significantly affect the overall communicative adequacy of the translation but remain underrepresented in purely metric-based assessments. Toral et al. (2018) found that automatic evaluation metrics are limited in their ability to measure referential relations that extend beyond the sentence-level. As they explain, "Our hypothesis is that referential relations that go beyond the sentence-level were ignored in the evaluation as its setup considered sentences in isolation (randomised). This probably resulted in the evaluation missing some errors by the MT system that might have been caused by its lack of inter-sentential contextual knowledge." (Toral et al., 2018:115). For this reason, it is essential to combine them with human assessment, particularly when employing frameworks such as MQM, which allow for a detailed categorisation of error types and their severity. Human evaluation enables a more nuanced understanding of the translation outputs, capturing aspects of meaning and context that automated metrics may overlook or underestimate.

In conclusion, the results of this study indicate that, although recent developments in machine translation have substantially enhanced output quality (especially regarding more recent texts), human intervention remains indispensable not only for the accurate interpretation of context and the optimisation of source texts but also for the assessment of translation quality. This corroborates the view that high-quality translation continues to emerge from the collaborative interaction between human expertise and computational processing.

CONCLUSIONS

This thesis examined the quality of the outputs from three different widespread engines, two Neural Machine Translation systems, namely DeepL and Google Translate and one Large Language Model, namely ChatGPT, to understand how they have performed on the translation of two different texts from Italian (L1) into English (L2): the first is from 2026, and is therefore characterised by simple syntactic structures and standardised, more translation-friendly language, whereas the second text is from 1956 and is distinguished by complex terminology and syntactic structures.

Translation quality was analysed both qualitatively, via two automatic metrics (COMET and TER), and quantitatively, via manual error classification through the MQM framework. The findings indicated that although these systems have undergone significant improvements over time, they still fail to produce fully accurate translations. Specifically, they produce contextual errors, namely instances in which the chosen lexical or structural equivalent is inappropriate because it fails to align with the specific contextual conditions arising both from the source text and from the rest of the target text, thereby undermining the overall comprehensibility and quality of the translation. This problem is particularly evident in the case of the 1956 text, where complex syntax and terminology lead to a comparatively high post-editing effort. The analysis also revealed that automatic evaluation metrics, especially COMET, exhibit limitations in penalising these types of errors. Consequently, the reliance on human expertise remains indispensable for quality assurance purposes through the application of post-editing techniques. In fact, the sentences underwent a full post-editing process (ISO:18587, 2017), which proved useful not only for conducting the analysis but also for generating the reference sentences employed in the automatic evaluation metrics, with the support of various online dictionaries and corpus platforms, too.

For this reason, the results confirm the crucial role of both post-editing and pre-editing as complementary strategies for improving translation quality. The former remains essential for correcting errors, especially those related to context and meaning. The latter can facilitate the processing of the source text by reducing ambiguity and simplifying lexical and syntactic structures, as noted for the 2026 article. However, as previously discussed, pre-editing may also entail drawbacks, as it can diminish the stylistic and linguistic richness of the source text in order to facilitate its processing by machine translation systems.

These findings reinforce the view that nowadays the role of the human translator is not diminished but rather transformed. Translators are increasingly required to act not only as language experts but also as evaluators and post-editors of machine-generated content, ensuring that the final output is accurate, coherent, and appropriate for the intended audience.

This study presents some limitations that should be acknowledged for a balanced interpretation of the results and for situating them within the broader, ever-evolving landscape of technology applied to translation. First, the analysis was based on a relatively small corpus of texts and focused on a specific high-resource language pair (Italian – English), which may limit the generalisability of the results. In addition, I produced the reference translations used for the evaluation. While the ideal methodological framework would involve native-level professional translators, the references provided here were specifically designed to align with the research objectives so as to ensure a solid foundation for the analysis in any case.

Furthermore, although the MQM framework enables a detailed and systematic classification of errors, it is inherently influenced by a degree of subjectivity, as the evaluation is conducted by human beings and it is considered to be "a time-consuming and resource-intensive process" (Rossi & Carré, 2022:53). Finally, it should be noted that automatic metrics such as COMET and TER typically operate at sentence-level rather than at document-level, which limits their ability to capture broader contextual dependencies and discourse-level phenomena that are crucial for a comprehensive assessment of translation quality (Toral et al., 2018).

Despite these limitations, the findings remain relevant for highlighting key tendencies in machine translation performance and for supporting the overall conclusions of this study.

Future research could build on this study by expanding the corpus to include a wider range of text types and domains, in order to verify whether the patterns observed in this study can be generalised across a wider range of translation scenarios. Further developments may also involve the investigation of machine translation outputs for additional language pairs, particularly those involving low-resource languages. Finally, greater attention could be dedicated to document-level evaluation, in order to better capture inter-sentential relations and contextual dependencies that are often overlooked by current metrics.

Overall, this study confirms that, despite the rapid progress of machine translation technologies, human expertise remains indispensable, not only in the translation process itself but also in the evaluation and improvement of machine-generated outputs. Rather than replacing the human translator, these technologies are redefining and reshaping their role, since the systems are not yet capable of producing error-free translations, regardless of whether the source text has undergone prior editing or has remained unedited.

BIBLIOGRAPHY

Bang, Y., Cahyawijaya, S., Lee, N., Dai, W., Su, J., Wilie, B., Lovenia, H., Ji, Z., Yu, T., Chung, C., Do, Q. V., Xu, Y., & Fung, P. (2023), A Multitask, Multilingual, Multimodal Evaluation of ChatGPT on Reasoning, Hallucination, and Interactivity. In *Proceedings of the 13th International Joint Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*, Association for Computational Linguistics, pp. 675-718.

Bernardini, S., Bouillon, P., Ciobanu, G., van Genabith, J., Hansen-Schirra, S., O'Brien, S., Steiner, E., Teich, E., (2020), Language service provision in the 21st century: Challenges, opportunities and educational perspectives for translation studies, in *Proceedings of Bologna process beyond 2020. Fundamental values of the EHEA*, Bologna, Bologna University Press, pp. 297-303.

Bowker, L., Buitrago Ciro, J., (2019), *Machine translation and global research*, Emerald Publishing, Bingley.

Campbell, S., (1998), *Translation into the second language*, Longman, New York.

Castilho, S., Moorkens, J., Gaspari, F., Calixto, I., Tinsley, J., Way, A., (2017), Is Neural Machine Translation the New State of the Art?, in *The Prague Bulletin of Mathematical Linguistics, No. 108*, 2017, pp. 109-120.

EMT, Eurpean Master's in Translation, Competence Framework, (2022), Dictatorate-General for Translation, Brussels. Available at https://commission.europa.eu/news-and-media/news/updated-version-emt-competence-framework-now-available-2022-10-21_en.

Holmes S., J., (1972), The name and nature of translation studies, in Venuti, L., *The Translation studies reader*, 2000, Routledge London and New York, pp. 172-185.

International Olympic Committee, (2026), *Olympic Charter*, https://www.olympics.com/ioc

Jiang, L., Yu, R., Zhao, Y., (2024), Theoretical perspectives and factors influencing machine translation use in L2 writing: A scoping review, in Ferris, D., Kibler, A. (eds.), *Journal of Second Language Writing*, 64, ELSEVIER.

Kachru, B. B., (1991), World Englishes and Applied Linguistics, in Tickoo, Makhan L. (eds.) *Languages & Standards: Issues, Attitudes, Case Studies*, Anthology Series 26, 1991, SEAMEO Regional Language Centre, Singapore, pp. 178-205.

Kenny D., (2022), Human and Machine Translation, in Kenny, D., (ed.) *Machine translation for everyone: Empowering users in the age of artificial intelligence*, (Translation and Multilingual Natural Language Processing 18), Language Science Press, Berlin, pp. 23-49.

Koby, G. S., Fields, P., Hague, D., Lommel, A., Melby, A., (2014), *Defining translation quality*. Revista Tradumàtica: Tecnologies de la Traducció, (12), 413-420.

Koponen, M., (2016), Is machine translation post-editing worth the effort? A survey of research into post-editing and effort, in *The Journal of Specialised Translation*, Issue 25, 2016.

Lommel, A., Uszkoreit, H., & Burchardt, A., (2014), Multidimensional Quality Metrics: A flexible system for assessing translation quality. *Traduire*, (231), 40-52.

Mason, I. (2006), On mutual accessibility of contextual assumptions in dialogue interpreting, *Journal of Pragmatics*, *38*(3), pp. 359–373. https://doi.org/10.1016/j.pragma.2005.06.015

Melby, A. K., (2020), Future of machine translation, Musings on Weaver's memo, in *The Routledge Handbook of Translation and Technology*, O'Hagan, M. (eds.), New York, Routledge, pp. 419-436.

Moorkens, J., Do Carmo, F., (2020), Differentiating editing, post-editing and revision, in Koponen, M., Mossop, B., S. Robert, I., & Scocchera, G. (Eds.), *Translation Revision and Post-editing: Industry Practices and Cognitive Processes (1st ed.)*, London and New York, Routledge, pp. 35-49.

Moorkens, J. (2022). Ethics and machine translation. In D. Kenny (Ed.), *Machine translation for everyone: Empowering users in the age of artificial intelligence*). Language Science Press, Berlin, pp. 119–140.

Munday, J., (2012), *Manuale di studi sulla traduzione*, seconda edizione, ed. italiana tradotta da Chiara Bucaria, Bononia University Press, Rastignano (BO).

Nitzke, J., Hansen-Schirra, S., (2021), *A short guide to post-editing*, (Translation and Multilingual Natural Language Processing 16), Language Science Press, Berlin.

O'Brien, S., (2011), Towards predicting post-editing productivity, in *Machine Translation 25(3),* pp. 197-215.

O'Brien, S., (2022), How to deal with errors in machine translation: Post editing, in Kenny., D. (ed.), *Machine translation for everyone: Empowering users in the age of artificial intelligence*, Language Science Press, Berlin, pp. 105-120.

PACTE, Building a translation competence model, (2003), in Alves, F. (ed.), *Triangulating translation: perspectives in process oriented research*, John Benjamins Publishing Company, Amsterdam, pp. 43-66.

Pavlović, N., (2008), Directionality in translation and interpreting practice, in Pym, A. and Perekrestenko, A. (eds.) *Translation Research Project 1. Tarragona: Intercultural Studies Group*, Universitat Rovira i Virgili, Tarragona, pp. 79-95.

Pérez-Ortiz, J. A., Forcada M. L., Sánchez-Martínez, F., (2022), How neural machine translation works, in Kenny, D. (ed.), *Machine translation for everyone: Empowering users in the age of artificial intelligence*, Language Science Press, Berlin, pp. 141-164.

Pokorn, N. K., (2005), *Challenging the Traditional Axioms – Translation into a non-mother tongue*, John Benjamins Publishing Company, Amsterdam/Philadelphia.

Rahaman, S., A., T., Anjum, N., Terano, H. J. R., Rahman, M., (2023), From ChatGPT-3 to GPT-4: A Significant Advancement in AI-Driven NLP Tools, in *Journal of Engineering and Emerging Technologies*, *2(1)*, pp. 1-11.

Rei, R., Stewart, C., Farinha, A. C., & Lavie, A., (2020), COMET: A neural framework for MT evaluation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*,), Association for Computational Linguistics, pp. 2685-2702.

Rico, C., Torrejòn, E., (2013), Skills and profile of the new role of the translator as MT post-editor, in *Tradùmatica technologies de la traducciò, 10*, pp. 167-178.

Robert, I. S., Remael, A., Ureel, J., (2017), Towards a model of translation revision competence, *The Interpreter and Translator Trainer 11(1)*, pp. 1-19.

Rossi, C., & Carré, A., (2022), How to choose a suitable neural machine translation solution: Evaluation of MT quality. In Dorothy Kenny (ed.), *Machine translation for everyone: Empowering users in the age of artificial intelligence*, 51–79. Language Science Press, Berlin.

Sánchez-Gijón, P., & Kenny, D., (2022). Selecting and preparing texts for machine translation: Pre-editing and writing for a global audience, in Kenny D. (ed.), *Machine translation for everyone: Empowering users in the age of artificial intelligence*, Language Science Press, Berlin, pp. 81-103.

Sizov, G., Öztürk, P., & Rothe, S. (2024). Analysing translation artifacts: A comparative study of LLMs, NMTs, and human translations, in *Proceedings of the Ninth Conference on Machine Translation*, Association for computational linguistics, pp. 1183–1199.

Snover, M., Dorr, B., Schwartz, R., Micciulla, L., & Makhoul, J., (2006), A study of translation edit rate with targeted human reference. In *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas (AMTA)*, pp. 223-231.

Stewart, D., (2020), English for tourism in the non-native English classroom: machine translation and corpora, in Ennis, M. J., Petrie, G. M. (eds.), *Teaching English for Tourism: Bridging research and practice*, Routledge, Abingdon UK, pp. 114-130.

Toral, A., Castilho, S., Hu, K., & Way, A. (2018), Attaining the unattainable? Reassessing claims of human parity in neural machine translation, in *Proceedings of the Third Conference on Machine Translation: Research Papers*, Association for Computational Linguistics, pp. 113-123. https://doi.org/10.18653/v1/W18-6312

Vu, T. L., Scherrer, Y., & Tiedemann, J. (2024), Context-aware machine translation with source coreference explanation. In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*, Association for Computational Linguistics, vol.12, pp. 856–874.

Zanettin, F., (2009), Corpus-based Translation Activities for Language Learners, in *The Interpreter and Translator Trainer, 3(2),* Routledge, London, pp. 209-224.

Online resources

Article of 1956
http://www.archiviolastampa.it/component/option,com_lastampa/task,search/mod,libera/action,viewer/Itemid,3/page,5/articleid,1586_02_1956_0022_0005_23771396/

Article of 2026 https://www.olympics.com/it/milano-cortina-2026/notizie/le-emozioni-della-cerimonia-di-apertura-di-milano-cortina-2026-i-valori-olimpici-in-una-celebrazione-diffusa

https://commission.europa.eu/get-involved/jobs-european-commission/working-eu-civil-service/translator-profile_it

https://www.ethnologue.com/insights/ethnologue200/

https://www.ethnologue.com/insights/most-spoken-language/

https://www.ibm.com/think/topics/deep-learning.

https://www.taus.net/company/about-us.

https://www.asd-ste100.org/.

https://qabiria.com/en/resources/blog/controlled-language.

https://unbabel.com/unbabel-launches-comet-blazing-new-trail-for-ultra-accurate-machine-translation

https://unbabel.com/research/comet/

https://themqm.org/

https://www.sketchengine.eu/

https://www.wordreference.com/

https://www.collinsdictionary.com/

https://www.merriam-webster.com/

https://www.ibm.com/think/topics/human-in-the-loop

https://www.english-corpora.org/now/

https://www.olympics.com/ioc

APPENDIX I

Article of 1956 and its translations

Stampa Sera, giovedì 26 – venerdì 27 gennaio 1956, anno X, numero 22

Aperte da mille atleti le Olimpiadi a Cortina

Ha assistito alla solenne cerimonia il Presidente della Repubblica on. Gronchi - Partecipano ai Giochi 32 Nazioni - Il pattinatore Caroli, portatore della fiaccola nell'ultima frazione, inciampa in un filo telefonico e cade - Giuliana Minuzzo ha pronunciato la formula del giuramento olimpico

Da uno dei nostri inviati. Cortina d'Ampezzo, giov. sera. I Giochi Olimpici si sono iniziati ufficialmente questa mattina alle 10.10, quando il Presidente della Repubblica nello Stadio del Ghiaccio, popolato di concorrenti e gremito di folla, si è alzato per pronunciare la formula di prammatica: «Dichiaro aperti i VII Giochi Olimpici Invernali di Cortina d'Ampezzo celebranti la XVI Olimpiade dell'era moderna». Poi la bandiera dai cerchi di cinque colori è salita lentamente sul pennone principale e l'artiglieria ha sparato tre salve di cannone che a lungo si sono ripercosse tra le gole dei monti incombenti. Innumerevoli persone hanno avuto modo di osservare lo svolgersi della cerimonia ai televisori, ma crediamo che soltanto coloro che sono stati presenti a Cortina abbiano potuto godere completamente del grandioso spettacolo, degno preludio alle competizioni. Vale la pena di farne una cronaca dettagliata. Si è cominciato alle 10.

Secondo una tabella di marcia rigorosamente stabilita (qualcuno ha mosso appunto di pignoleria agli organizzatori, ma noi crediamo che anche i Giochi devono essere presi molto sul serio quando si fanno al cospetto del mondo), le delegazioni alloggiate fuori Cortina, cioè Russia, Olanda e Corea del Sud, sono partite dalle loro sedi per trovarsi sul piazzale della stazione con pullman forniti dal Comitato. Qui sono affluite alle 10.50 tutte le altre ventinove delegazioni alloggiate in città. Non è stato difficile incolonnarle: in testa la Grecia, perché questo onore spetta al popolo che fondò le Olimpiadi, poi tutte le altre in ordine alfabetico fino all'Italia, sistemata in coda per il fatto di essere padrona di casa. Elenchiamo fin d'ora i 38 Paesi partecipanti: Grecia, Australia, Austria, Belgio, Bolivia, Bulgaria, Canada, Cecoslovacchia, Cile, Corea, Finlandia, Francia, Germania, Giappone, Gran Bretagna, Iran, Islanda, Jugoslavia, Libano, Liechtenstein, Norvegia, Olanda, Polonia, Rumenia, Spagna, Stati Uniti, Svezia, Svizzera, Turchia, Ungheria, U.R.S.S. e Italia. In totale 945 atleti e 94 accompagnatori. Questi ultimi sembrano troppi; facciamo notare che sono invece pochi rispetto alla reale quantità degli «ufficiali» scesi a Cortina al seguito dei concorrenti. Qualche esempio preso a caso: la Corea del Sud, che evidentemente non bada a spese, ha mandato 4 atleti e 3 accompagnatori; il Giappone rispettivamente 10 e 10 (si vede che l'Italia attira enormemente gli orientali); il Libano 3 e 5, la Russia 68 e 44, l'Italia 78 e 34.

Di fronte a queste cifre il Comitato organizzatore si è preoccupato ed ha deciso misure drastiche; il numero dei cosiddetti “ufficiali” di ogni Paese non può essere superiore, durante la sfilata, al 10 % dei rispettivi concorrenti. Molti accompagnatori sono quindi rimasti in tribuna o in albergo, ma lo spettacolo ci ha guadagnato in snellezza ed eleganza. Così la colonna variopinta, preceduta da un gruppo di 25 alfieri (allievi dell'Istituto superiore di educazione fisica di Roma) reggenti le bandiere olimpiche, è sfilata per le vie di Cortina ed ha raggiunto il piazzale esterno dello Stadio del Ghiaccio. Intanto sono le 11. Arriva il Presidente della Repubblica. Sono con lui il presidente del Senato, Merzagora, e il presidente della Camera, Leone, il ministro dei Trasporti, Angelini, in rappresentanza del Governo, l’on. Negrari, presidente del gruppo parlamentare dello sport. Li ricevono il presidente del Comitato Internazionale Olimpico, che è l'americano Avery Brundage, e il presidente del Comitato organizzatore italiano, conte Paolo Thaon di Revel. La fanfara suona l'attenti e l’on. Gronchi passa in rivista la compagnia d'onore con bandiera del 7° reggimento alpini; poi le autorità entrano nello stadio, le bandiere delle 38 nazioni partecipanti salgono lentamente sui pennoni. La sfilata delle squadre ha inizio alle 11,45. Così dice il programma, e controlliamo i cronometri. Sono soltanto le 11.42 e qualche secondo: veramente strano, questo anticipo.

I meticolosi organizzatori rimettono le cose a posto facendo rallentare la marcia di quel tanto che basta per andare di nuovo d'accordo con la tabella-orario. (Dimenticavamo di annotare un'altra sottigliezza: fra l'atleta che regge il cartello con il nome del suo Paese e l'alfiere con la bandiera devono passare metri tre; tra l'alfiere e la delegazione metri cinque. Non abbiamo il metro, non abbiamo potuto controllare se le misure corrispondono agli ordini). Le squadre sono disposte su file di tre; prima gli "ufficiali", poi le rappresentanze femminili, poi le rappresentanze maschili. Una parata di gioventù magnifica. Ecco la Grecia e l'Australia, ecco la fitta rappresentanza austriaca che si appresta a raccogliere i massimi onori nelle prove di discesa, ecco i dodici belgi, poi un gruppetto di due sole persone: la Bolivia, un accompagnatore e un atleta che farà lo slalom gigante con infinitesime possibilità di successo. La prima squadra numerosa è il Canada, con 37 concorrenti, ma la Cecoslovacchia ne ha 41 e batte anche la Finlandia, che pure ne schiera 36. Verso il termine tre squadroni: i russi, gli americani, gli italiani, che sono i gruppi più numerosi. All'apparire dei nostri l'urlo della folla diviene tempesta: non sappiamo se gli amplificatori degli apparecchi televisivi hanno potuto registrarlo con esattezza. Guida i nostri il saltatore Nilo Zandanel.

Al suono della marcia olimpica la colonna si dispone sulla lucida pista dello stadio a formare i tre lati d'un rettangolo: il quarto è aperto di fronte alla tribuna d'onore. La fanfara tace, il conte di Revel con brevi parole invita il Capo dello Stato a proclamare l'inizio delle Olimpiadi. Nel silenzio assoluto l'on. Gronchi si alza e pronuncia la formula. Ma già l'attenzione della folla è volta ad altro. Sta arrivando la fiaccola di Olimpia. Zeno Colò l'ha portata per la prima frazione detta staffetta dal rifugio «Duca d'Aosta», poi la face è passata al fondista Menardi, al discesista Ghedina, al fondista Colli, al conquistatore del K2 Lacedelli, a un altro fondista che si chiama ancora Colli, finalmente al pattinatore Caroli. Per permettere l'arrivo allo stadio di questi staffettisti molti alpini hanno lavorato duramente per tutta la notte affinché almeno una stretta striscia di neve si dilungasse dalla boscaglia delle Tofane fino alla pista della cerimonia. Ed ecco Caroli entrare nello stadio, accompagnato dall'urlo della moltitudine. La sorte, l'emozione e un filo telefonico gli giocano però una trista beffa. Il campione italiano della velocità sui pattini, primatista dei 500 m. 5000 e 10.000, a metà del giro d'onore inciampa nell'ostacolo imprevisto e cade sul fianco, quasi davanti alla tribuna delle autorità. Nella disavventura ha la presenza di spirito di tenere alta la fiaccola, e questa pertanto non si spegne.

Pochi minuti dopo Caroli sale sul tripode e accende il fuoco olimpica. Ora una figurina lo segue sul palco: la nostra Giuliana Minuzzo, medaglia di bronzo ad Oslo, campionessa di discesa. Forse è più emozionata che alla partenza di una gara di grande impegno. Afferra con la destra un lembo della bandiera italiana e mentre gli altri vessilli si piegano sino a sfiorare la pista, pronuncia la formula del rito: "Noi giuriamo di partecipare ai Giochi Olimpici come concorrenti leali rispettosi dei regolamenti che li reggono e desiderosi di gareggiare con spirito cavalleresco por l'onore del nostro Paese e la gloria dello sport». Ci è parso di avvertire un tremito nella voce di Giuliana. Il marito, che è in tribuna al nostro fianco, ha gli occhi pieni di lacrime. Si levano solenni le note di un lento inno di montagna, mentre il Presidente della Repubblica e le autorità lasciano lo stadio. Gli atleti iniziano il movimento per uscire dalla pista. La cerimonia è finita. Da questo momento la migliore gioventù del mondo si batte pacificamente per un primato di gloria sportiva.

Carlo Moriondo

The Olympics in Cortina opened by a thousand athletes (DEEPL)

The President of the Republic, Mr Gronchi, attended the solemn ceremony – 32 nations are participating in the Games – Skater Caroli, torchbearer in the last leg, tripped over a telephone wire and fell – Giuliana Minuzzo recited the Olympic oath.

From one of our correspondents. Cortina d'Ampezzo, Thursday evening.

The Olympic Games officially began this morning at 10.10 a.m., when the President of the Republic stood up in the Ice Stadium, filled with competitors and packed with crowds, to pronounce the customary formula: “I declare open the VII Winter Olympic Games in Cortina d'Ampezzo, celebrating the XVI Olympiad of the modern era”. Then the flag with the five coloured circles slowly rose on the main flagpole and the artillery fired three cannon salutes that echoed for a long time in the gorges of the looming mountains. Countless people were able to watch the ceremony on television, but we believe that only those who were present in Cortina were able to fully enjoy the magnificent spectacle, a worthy prelude to the competitions. It is worth giving a detailed account of it. It began at 10 o'clock.

According to a strictly established schedule (someone criticised the organisers for being too pedantic, but we believe that even the Games must be taken very seriously when they are held in front of the whole world), the delegations staying outside Cortina, namely Russia, the Netherlands and South Korea, left their accommodation to gather in the station square with coaches provided by the Committee. All the other 29 delegations staying in the city arrived here at 10.50 a.m. It was not difficult to line them up: Greece at the head, because this honour belongs to the people who founded the Olympics, then all the others in alphabetical order up to Italy, placed at the end because it was the host country. Here is a list of the 38 participating countries: Greece, Australia, Austria, Belgium, Bolivia, Bulgaria, Canada, Czechoslovakia, Chile, Korea, Finland, France, Germany, Japan, Great Britain, Iran, Iceland, Yugoslavia, Lebanon, Liechtenstein, Norway, Holland, Poland, Romania, Spain, United States, Sweden, Switzerland, Turkey, Hungary, USSR and Italy. A total of 945 athletes and 94 accompanying persons. The latter seem too many; we would point out that they are actually few compared to the real number of “officials” who came to Cortina accompanying the competitors. A few random examples: South Korea, which clearly spared no expense, sent four athletes and three accompanying persons; Japan sent ten and ten respectively (it is clear that Italy is a huge attraction for Asians); Lebanon 3 and 5, Russia 68 and 44, Italy 78 and 34.

Faced with these figures, the organising committee became concerned and decided on drastic measures; the number of so-called “officials” from each country could not exceed 10% of the respective competitors during the parade. Many accompanying persons therefore remained in the stands or in their hotels, but the show gained in streamlined elegance. Thus, the colourful column, preceded by a group of 25 flag bearers (students from the Higher Institute of Physical Education in Rome) carrying the Olympic flags, paraded through the streets of Cortina and reached the square outside the Ice Stadium. Meanwhile, it is 11 o'clock. The President of the Republic arrives. He is accompanied by the President of the Senate, Merzagora, and the President of the Chamber of Deputies, Leone, the Minister of Transport, Angelini, representing the Government, and the Honourable Negrari, President of the Parliamentary Sports Group. They are received by the President of the International Olympic Committee, Avery Brundage of the United States, and the President of the Italian Organising Committee, Count Paolo Thaon di Revel. The fanfare sounds the attention call and Hon. Gronchi reviews the honour guard with the flag of the 7th Alpine Regiment; then the authorities enter the stadium, and the flags of the 38 participating nations are slowly raised on the flagpoles. The parade of teams begins at 11.45 a.m., according to the programme, and we check our stopwatches. It is only 11.42 a.m. and a few seconds: this early start is really strange.

The meticulous organisers put things back in order by slowing down the march just enough to get back on schedule. (We forgot to note another subtle detail: there must be three metres between the athlete holding the sign with the name of his country and the flag bearer; five metres between the flag bearer and the delegation. We don't have a metre, so we couldn't check whether the measurements correspond to the orders). The teams are arranged in rows of three: first the “officials”, then the women's teams, then the men's teams. A magnificent parade of youth. Here are Greece and Australia, here is the large Austrian delegation preparing to reap the highest honours in the downhill events, here are the twelve Belgians, then a small group of only two people: Bolivia, an accompanying person and an athlete who will compete in the giant slalom with infinitesimal chances of success. The first large team is Canada, with 37 competitors, but Czechoslovakia has 41 and even beats Finland, which also fields 36. Towards the end, there are three large teams: the Russians, the Americans and the Italians, which are the largest groups. When our team appeared, the crowd's roar became a storm: we do not know if the television amplifiers were able to record it accurately. Our team was led by the jumper Nilo Zandanel.

To the sound of the Olympic march, the column lines up on the shiny track of the stadium to form three sides of a rectangle: the fourth side is open in front of the VIP stand. The fanfare falls silent, and Count Revel briefly invites the Head of State to proclaim the start of the Olympics. In absolute silence, the Honourable Gronchi stands up and pronounces the formula. But the crowd's attention is already focused on something else. The Olympic torch is arriving. Zeno Colò carried it for the first leg, known as the relay, from the “Duca d'Aosta” refuge, then it was passed to cross-country skier Menardi, downhill skier Ghedina, cross-country skier Colli, K2 conqueror Lacedelli, another cross-country skier also named Colli, and finally to skater Caroli. To enable these relay runners to reach the stadium, many Alpine troops worked hard throughout the night to ensure that at least a narrow strip of snow stretched from the Tofane woods to the ceremony track. And so Caroli entered the stadium, accompanied by the cheers of the crowd. However, fate, emotion and a telephone wire played a cruel joke on him. The Italian speed skating champion, record holder in the 500 m, 5000 m and 10,000 m, stumbled over the unexpected obstacle halfway through his lap of honour and fell on his side, almost in front of the VIP stand. In his misfortune, he has the presence of mind to hold the torch high, and so it does not go out.

A few minutes later, Caroli climbs onto the tripod and lights the Olympic flame. Now a figure follows him onto the stage: our Giuliana Minuzzo, bronze medallist in Oslo, downhill champion. Perhaps she is more excited than at the start of a major competition. She grabs a corner of the Italian flag with her right hand and, while the other flags are lowered to touch the track, she recites the ritual formula: “We swear to participate in the Olympic Games as fair competitors, respecting the rules that govern them and eager to compete in a spirit of chivalry for the honour of our country and the glory of sport”. We thought we detected a tremor in Giuliana's voice. Her husband, who is in the stands next to us, has tears in his eyes. The solemn notes of a slow mountain anthem rise as the President of the Republic and the authorities leave the stadium. The athletes begin to leave the track. The ceremony is over. From this moment on, the best young people in the world will compete peacefully for sporting glory.

The Olympics opened in Cortina with a thousand athletes (GOOGLE TRANSLATE)

The President of the Republic, Hon. Gronchi, attended the solemn ceremony. 32 nations are participating in the Games. Figure skater Caroli, torch bearer in the final leg, trips on a telephone wire and falls. Giuliana Minuzzo recited the Olympic oath.

From one of our correspondents. Cortina d'Ampezzo, Thursday evening. The Olympic Games officially began this morning at 10:10, when the President of the Republic stood in the packed Ice Stadium to pronounce the formal oath: "I declare open the VII Olympic Winter Games of Cortina d'Ampezzo, celebrating the XVI Olympiad of the modern era." Then the five-colored circled flag slowly rose to the main mast and the artillery fired three cannon salutes that echoed for a long time through the gorges of the looming mountains. Countless people watched the ceremony unfold on television, but we believe only those present in Cortina were able to fully enjoy the grandiose spectacle, a worthy prelude to the competitions. It's worth a detailed account. It began at 10:00.

According to a strictly established timetable (some have criticized the organizers for being fussy, but we believe that even the Games must be taken very seriously when held in front of the world), the delegations staying outside Cortina—Russia, the Netherlands, and South Korea—departed from their bases to meet in the station square on buses provided by the Committee. At 10:50 a.m., all twenty-nine other delegations staying in the city converged there. It wasn't difficult to line them up: Greece led the way, because this honor belongs to the people who founded the Olympics, followed by all the others in alphabetical order, ending with Italy, placed at the end of the queue because it was the host country. Let's list the 38 participating countries: Greece, Australia, Austria, Belgium, Bolivia, Bulgaria, Canada, Czechoslovakia, Chile, Korea, Finland, France, Germany, Japan, Great Britain, Iran, Iceland, Yugoslavia, Lebanon, Liechtenstein, Norway, the Netherlands, Poland, Romania, Spain, Sweden, Switzerland, Turkey, Hungary, the USSR, and Italy. A total of 945 athletes and 94 accompanying persons. The latter number may seem excessive; it should be noted that they are actually few compared to the actual number of "officials" who descended on Cortina to accompany the competitors. A few random examples: South Korea, evidently sparing no expense, sent four athletes and three accompanying persons; Japan 10 and 10 respectively (it's clear that Italy is a huge draw for Easterners); Lebanon 3 and 5, Russia 68 and 44, and Italy 78 and 34.

Faced with these numbers, the organizing committee became concerned and decided on drastic measures; the number of so-called "officials" from each country during the parade could not exceed 10% of their respective competitors. Many of their companions therefore remained in the stands or at their hotels, but the spectacle was enhanced in both elegance and speed. Thus, the colorful convoy, preceded by a group of 25 standard-bearers (students from the Higher Institute of Physical Education in Rome) holding Olympic flags, paraded through the streets of Cortina and reached the square outside the Ice Stadium. Meanwhile, it was 11:00 a.m. The President of the Republic arrived. With him were the President of the Senate, Merzagora, and the Speaker of the Chamber, Leone, the Minister of Transport, Angelini, representing the Government, and the Hon. Negrari, president of the parliamentary sports group. They are received by the president of the International Olympic Committee, American Avery Brundage, and the president of the Italian Organizing Committee, Count Paolo Thaon di Revel. The brass band sounds attention, and the Honorable Gronchi reviews the honor company carrying the flag of the 7th Alpine Regiment; then the dignitaries enter the stadium, and the flags of the 38 participating nations slowly rise to the flagpoles. The team parade begins at 11:45. That's what the program says, and we check our stopwatches. It's only 11:42 and a few seconds: truly strange, this early.

The meticulous organizers put things right, slowing the pace just enough to get back on track. (We forgot to note another subtlety: there must be three meters between the athlete holding the sign with his country's name and the standard-bearer with the flag; five meters between the standard-bearer and the delegation. We don't have the tape measure; we couldn't check whether the measurements matched the orders.) The teams are lined up in rows of three; first the "officials," then the women's representatives, then the men's representatives. A magnificent parade of youth. Here are Greece and Australia, here is the dense Austrian contingent preparing to reap the highest honors in the downhill events, here are the twelve Belgians, then a small group of just two: Bolivia, a companion, and an athlete who will compete in the giant slalom with an infinitesimal chance of success. The first largest team is Canada, with 37 competitors, but Czechoslovakia has 41 and also beats Finland, which also fields 36. Towards the end, three teams emerge: the Russians, the Americans, and the Italians, who are the largest groups. When our team appears, the roar of the crowd escalates into a storm: we don't know if the television amplifiers were able to accurately record it. Leading our team is high jumper Nilo Zandanel.

At the sound of the Olympic march, the column lines up on the stadium's polished track to form three sides of a rectangle: the fourth is open in front of the grandstand. The fanfare falls silent, and Count Revel briefly invites the Head of State to proclaim the start of the Olympics. In the absolute silence, the Honorable Gronchi stands and pronounces the formula. But the crowd's attention is already turned elsewhere. The Olympic torch is arriving. Zeno Colò carried it for the first leg, called the relay, from the "Duca d'Aosta" refuge. Then the torch is passed to cross-country skier Menardi, downhill skier Ghedina, cross-country skier Colli, K2 conqueror Lacedelli, another cross-country skier also named Colli, and finally to figure skater Caroli. To enable these relay runners to reach the stadium, many Alpine troops worked hard throughout the night to ensure at least a narrow strip of snow stretched from the Tofane woods to the ceremonial track. And then Caroli entered the stadium, accompanied by the roar of the crowd. However, fate, emotion, and a telephone wire played a cruel trick on him. The Italian speed skating champion, record holder in the 500m, 5000m, and 10,000m, tripped over an unexpected obstacle midway through his lap of honor and fell on his side, almost in front of the officials' stands. In his misfortune, he had the presence of mind to hold the torch high, and so it remained.

A few minutes later, Caroli climbs onto the tripod and lights the Olympic flame. Now a figurine follows him onto the stage: our own Giuliana Minuzzo, bronze medalist in Oslo and downhill champion. Perhaps she's more excited than she is at the start of a challenging race. He grasps a corner of the Italian flag in his right hand, and as the other flags fold until they touch the track, he pronounces the ritual formula: "We swear to participate in the Olympic Games as fair competitors, respectful of the regulations that govern them and eager to compete with a spirit of chivalry for the honor of our country and the glory of sport." We thought we detected a tremor in Giuliana's voice. Her husband, who is in the stands beside us, has tears in his eyes. The solemn strains of a slow mountain hymn rise, while the President of the Republic and the authorities leave the stadium. The athletes begin their movement to exit the track. The ceremony is over. From this moment, the finest youth in the world will fight peacefully for a record of sporting glory.

The Cortina Olympics Opened by a Thousand Athletes (CHATGPT)
President of the Republic Hon. Gronchi Attends the Solemn Ceremony – 32 Nations Take Part in the Games – Skater Caroli, Torchbearer in the Final Leg, Trips Over a Telephone Wire and Falls – Giuliana Minuzzo Recites the Olympic Oath

From one of our correspondents. Cortina d’Ampezzo, Thursday evening.

The Olympic Games officially began this morning at 10:10, when the President of the Republic, in the Ice Stadium filled with competitors and packed with spectators, rose to pronounce the customary formula: “I declare open the VII Winter Olympic Games of Cortina d’Ampezzo, celebrating the XVI Olympiad of the modern era.” Then the flag with the five-colored rings slowly rose on the main flagpole, and the artillery fired three cannon salvos that echoed for a long time among the surrounding mountain gorges. Countless people were able to follow the ceremony on television screens, but we believe that only those who were present in Cortina could fully enjoy the magnificent spectacle, a worthy prelude to the competitions. It is worth giving a detailed account of it. Proceedings began at 10 o’clock.

According to a strictly established timetable (some have accused the organizers of excessive fussiness, but we believe that even the Games must be taken very seriously when staged before the eyes of the world), the delegations lodged outside Cortina — namely Russia, the Netherlands, and South Korea — departed from their quarters to gather at the station square on buses provided by the Committee. Here, at 10:50, all the other twenty-nine delegations staying in town arrived. It was not difficult to line them up: Greece at the head, since this honor belongs to the people who founded the Olympics, followed by all the others in alphabetical order, with Italy at the rear as the host nation. We list at once the 38 participating countries: Greece, Australia, Austria, Belgium, Bolivia, Bulgaria, Canada, Czechoslovakia, Chile, Korea, Finland, France, Germany, Japan, Great Britain, Iran, Iceland, Yugoslavia, Lebanon, Liechtenstein, Norway, the Netherlands, Poland, Romania, Spain, the United States, Sweden, Switzerland, Turkey, Hungary, the U.S.S.R., and Italy. In total, 945 athletes and 94 accompanying officials. The latter seem too many; we note, however, that they are actually few compared with the real number of "officials" who have come to Cortina in the athletes' wake. A few random examples: South Korea, which clearly spares no expense, sent 4 athletes and 3 officials; Japan, respectively 10 and 10 (Italy evidently exerts enormous attraction on Easterners); Lebanon 3 and 5; Russia 68 and 44; Italy 78 and 34.

Faced with these figures, the Organizing Committee grew concerned and decided on drastic measures: during the parade, the number of so-called "officials" for each country could not exceed 10 percent of the number of competitors. Many officials therefore remained in the stands or at their hotels, but the spectacle gained in slimness and elegance. Thus the colorful column, preceded by a group of 25 standard-bearers (students of the Higher Institute of Physical Education of Rome) carrying the Olympic flags, paraded through the streets of Cortina and reached the open area in front of the Ice Stadium. Meanwhile it is 11 o'clock. The President of the Republic arrives. With him are the President of the Senate, Merzagora, the President of the Chamber of Deputies, Leone, the Minister of Transport, Angelini, representing the Government, and Hon. Negrari, president of the parliamentary sports group. They are received by the President of the International Olympic Committee, the American Avery Brundage, and by the President of the Italian Organizing Committee, Count Paolo Thaon di Revel. The fanfare plays the call to attention and Hon. Gronchi reviews the honor guard with the flag of the 7th Alpine Regiment; then the authorities enter the stadium, and the flags of the 38 participating nations slowly rise on their poles. The parade of teams is scheduled to begin at 11:45. So says the program, and we check our watches. It is only 11:42 and a few seconds: truly strange, this early start.

The meticulous organizers put things right by slowing the march just enough to bring it back into line with the timetable. (We forgot to note another subtlety: between the athlete carrying the sign with his country's name and the flag-bearer there must be three meters; between the flag-bearer and the delegation, five meters. We did not have a tape measure and therefore could not check whether the distances complied with orders.) The teams are arranged in rows of three: first the "officials," then the women's representatives, then the men's. A magnificent parade of youth. Here come Greece and Australia; here is the strong Austrian contingent, preparing to reap the highest honors in the downhill events; here are the twelve Belgians; then a tiny group of just two people: Bolivia, one official and one athlete who will compete in the giant slalom with infinitesimal chances of success. The first large team is Canada, with 37 competitors, but Czechoslovakia has 41 and even surpasses Finland, which fields 36. Toward the end come three large squads: the Russians, the Americans, and the Italians, the largest groups of all. At the appearance of our own team, the roar of the crowd becomes a tempest; we do not know whether television amplifiers were able to record it accurately. Leading our athletes is the ski jumper Nilo Zandanel.

To the sound of the Olympic march, the column takes position on the gleaming ice of the stadium, forming three sides of a rectangle; the fourth is open toward the grandstand of honor. The fanfare falls silent, and Count di Revel, with brief words, invites the Head of State to proclaim the opening of the Olympics. In absolute silence, Hon. Gronchi rises and pronounces the formula. But already the crowd's attention has turned elsewhere. The torch from Olympia is arriving. Zeno Colò carried it for the first leg of the relay from the "Duca d'Aosta" refuge; then the flame passed to the cross-country skier Menardi, the downhill racer Ghedina, the cross-country skier Colli, the conqueror of K2 Lacedelli, another cross-country skier also named Colli, and finally to the skater Caroli. To allow these relay runners to reach the stadium, many Alpine troops worked hard throughout the night so that at least a narrow strip of snow could extend from the woods of the Tofane down to the ceremonial track.

And now Caroli enters the stadium, accompanied by the roar of the multitude. Fate, emotion, and a telephone wire, however, play him a cruel trick. The Italian speed-skating champion, record-holder in the 500, 5,000, and 10,000 meters, halfway through his lap of honor trips over the unforeseen obstacle and falls onto his side, almost in front of the authorities' stand. In his mishap he has the presence of mind to keep the torch raised, and thus it does not go out.

A few minutes later Caroli climbs the tripod and lights the Olympic flame. Now a small figure follows him onto the platform: our Giuliana Minuzzo, bronze medalist at Oslo, downhill champion. Perhaps she is more emotional than at the start of a highly demanding race. With her right hand she grasps a fold of the Italian flag, and while the other banners bow until they nearly touch the ice, she pronounces the ritual formula: "We swear that we will take part in the Olympic Games as loyal competitors, respecting the rules that govern them and eager to compete in a chivalrous spirit, for the honor of our country and the glory of sport." It seemed to us that we detected a tremor in Giuliana's voice. Her husband, who is in the stands beside us, has tears in his eyes. The solemn notes of a slow mountain hymn rise, as the President of the Republic and the authorities leave the stadium. The athletes begin to move off the ice. The ceremony is over. From this moment, the finest youth of the world will compete peacefully for a title of sporting glory.

Article of 2026 and its translations

Le emozioni della Cerimonia di Apertura di Milano Cortina 2026: i valori Olimpici in una celebrazione diffusa

Sotto il segno dell'Armonia, i XXV Giochi Olimpici Invernali hanno connesso in tempo reale lo Stadio San Siro di Milano con gli altri territori parte integrante del grande evento sportivo.

La Cerimonia di Apertura di Milano Cortina 2026 ha segnato una svolta storica: per la prima volta nella storia dei Giochi, la Cerimonia ha preso forma in modalità diffusa, coinvolgendo Milano, Cortina, Livigno, Predazzo e Antholz/Anterselva in un unico racconto orchestrato in tempo reale.

Il concept creativo, intitolato "Armonia", ha rappresentato la capacità di far convivere elementi differenti, metterli insieme: città e montagna, uomo e natura.

Per rispettare il principio di sostenibilità e ridurre al minimo gli spostamenti, gli atleti e le atlete non sono stati raggruppati a Milano, ma hanno sfilato nelle quattro sedi più vicine ai rispettivi Villaggi Olimpici: Milano (Stadio San Siro), Cortina d'Ampezzo (Piazza Angelo Dibona), Livigno e Predazzo.

Durante la serata, le esibizioni di artisti, artiste, performer e il pieno di momenti clou: da Mariah Carey che ha cantato Nel Blu, dipinto di Blu di Domenico Modugno, seguito da uno dei suoi brani più iconici, Nothing Is Impossible a le esibizioni di Matilda De Angelis, Pierfrancesco Favino e i momenti di spettacolo live guidati da Sabrina Impacciatore, Ghali e Charlize Theron, questi ultimi al centro del quadro dedicato alla Tregua Olimpica e alla riflessione sul rifiuto della guerra.

Ecco alcuni dei momenti da ricordare.

Le emozioni della serata

1. L'Inno Nazionale e l'abbraccio del Tricolore

Il momento: il protocollo istituzionale si è svolto simultaneamente a Milano e Cortina. L'Inno Nazionale è stato eseguito da Laura Pausini allo Stadio San Siro, accompagnata da un suggestivo Coro di Montagna a Cortina d'Ampezzo. La top model Vittoria Ceretti, che ha sfilato insieme a un gruppo di modelle che hanno indossato creazioni disegnate da Giorgio Armani, è stata scelta per rappresentare lo spirito creativo e contemporaneo della città ha portato la Bandiera Nazionale fino al palco protocollare, dove il vessillo è stato affidato al Corpo dei Corazzieri.

In simultanea, a Cortina, un gruppo di portabandiera, espressione della tradizione sportiva, Fulvio Valbusa, Giorgio Di Centa, Pietro Piller Cottrer, Cristian Zorzi, compie lo stesso gesto, consegnando la Bandiera Italiana al Corpo dei Carabinieri.

2. La Parata degli Atleti: un'unica sfilata per i territori

Il momento: per la prima volta nella storia Olimpica, la Parata degli Atleti si è svolta in forma diffusa, collegando simultaneamente Milano (San Siro), Livigno (Snowboard Park), Predazzo (Stadio del Salto) e Cortina (Centro città). Le 92 delegazioni, partendo dalla Grecia e chiudendo con l'Italia, hanno sfilato nei siti più vicini alle proprie sedi di gara, trasformando la geografia italiana in un unico palcoscenico condiviso.

Scelta simbolica dell'Italia che, per rappresentare l'unità del Paese, ha schierato eccezionalmente quattro portabandiera: due a Milano. Arianna Fontana e Federico Pellegrino e due a Cortina, Federica Brignone e Amos Mosaner.

3. Viaggio nel tempo: dai gesti di Munari alla grande storia

Il momento: un segmento artistico che ha ripercorso 100 anni di Olimpiadi Invernali, con una performance di Brenda Lodigiani dedicata alla lingua dei gesti italiani raccontata da Bruno Munari.

I discorsi e il protocollo

Giovanni Malagò (Presidente del Comitato Organizzatore): "Stasera l'Italia apre le braccia al mondo - afferma Malagò, che ricorda i precedenti Giochi ospitati dal nostro Paese – Siamo pronti a scrivere di nuovo la storia delle Olimpiadi".

Kirsty Coventry (Presidente del Comitato Olimpico Internazionale): "Ce l'avete fatta, siate fieri di essere arrivati fin qui. Ora date il meglio di voi, divertitevi, assaporate ogni momento. Ci regalerete qualcosa di speciale, ci farete sognare, ci mostrerete che la forza non è solo una questione di vittorie ma anche di coraggio, empatia e cuore. Ci insegnerete a rialzarci, non importa quanto pesante sarà stata la caduta".

Sergio Mattarella (Presidente della Repubblica italiana): dalla tribuna d'onore, il Presidente della Repubblica Italiana, Sergio Mattarella, ha dichiarato ufficialmente aperti i Giochi Olimpici Invernali Milano Cortina 2026.

4. Il Viaggio della Fiamma Olimpica: una staffetta collettiva

Il momento: sugli schermi è scivolato il racconto visivo del viaggio della Fiamma, partita da Olimpia per attraversare i paesaggi italiani fino a unire Milano, Cortina e tutte le sedi dei Giochi.

All'interno dello stadio San Siro, la Torcia "Essential" è stata protagonista di una staffetta simbolica: prima portata da tre tedofore, tra cui Paola Egonu, poi affidata a due successivi gruppi di tre atleti italiani per l'uscita dallo stadio accompagnati dalla performance di Andrea Bocelli.

5. La Bandiera Olimpica: un ponte tra generazioni e territori

Il momento: la Bandiera Olimpica ha fatto il suo ingresso solenne seguendo la logica diffusa dell'evento. Portata da dieci leggende dello sport e personalità di spicco (otto a Milano e due a Cortina), il vessillo con i Cinque Cerchi ha attraversato i luoghi simbolo della Cerimonia.

6. L'accensione dei Bracieri

Il momento: gli ultimi tre tedofori, i campioni Olimpici di sci alpino Alberto Tomba e Deborah Compagnoni, hanno acceso contemporaneamente il Braciere all'Arco della Pace a Milano, mentre la campionessa Olimpica di sci alpino Sofia Goggia ha acceso il Braciere in Piazza Angelo Dibona a Cortina.

Così la Fiamma Olimpica ha unito la dimensione urbana di Milano con la maestosità delle Dolomiti patrimonio UNESCO.

The excitement of the Milan Cortina 2026 Opening Ceremony: Olympic values in a widespread celebration (DEEPL)

Under the banner of Harmony, the XXV Olympic Winter Games connected Milan's San Siro Stadium in real time with other areas that were an integral part of this major sporting event.

The Milan Cortina 2026 Opening Ceremony marked a historic turning point: for the first time in the history of the Games, the Ceremony took shape in a widespread manner, involving Milan, Cortina, Livigno, Predazzo and Antholz/Anterselva in a single story orchestrated in real time.

The creative concept, entitled "Harmony", represented the ability to bring different elements together: city and mountains, man and nature.

In order to respect the principle of sustainability and minimise travel, the athletes were not grouped together in Milan but paraded in the four locations closest to their respective Olympic Villages: Milan (San Siro Stadium), Cortina d'Ampezzo (Piazza Angelo Dibona), Livigno and Predazzo.

The evening featured performances by artists and performers and was full of highlights: from Mariah Carey singing Nel Blu, dipinto di Blu by Domenico Modugno, followed by one of her most iconic songs, Nothing Is Impossible, to performances by Matilda De Angelis. Pierfrancesco Favino and live performances led by Sabrina Impacciatore, Ghali and Charlize Theron, the latter two at the centre of the segment dedicated to the Olympic Truce and reflections on the rejection of war. Here are some of the moments to remember.

The emotions of the evening

1. The National Anthem and the embrace of the Tricolour

The moment: the institutional protocol took place simultaneously in Milan and Cortina. The National Anthem was performed by Laura Pausini at the San Siro Stadium, accompanied by an evocative Mountain Choir in Cortina d'Ampezzo. Top model Vittoria Ceretti, who walked the runway alongside a group of models wearing creations designed by Giorgio Armani, was chosen to represent the creative and contemporary spirit of the city. She carried the National Flag to the protocol stage, where the banner was entrusted to the Corps of the Corazzieri.

Simultaneously, in Cortina, a group of flag bearers representing sporting tradition, Fulvio Valbusa, Giorgio Di Centa, Pietro Piller Cottrer and Cristian Zorzi, performed the same gesture, handing over the Italian flag to the Carabinieri Corps.

2. The Athletes' Parade: a single parade for the territories

The moment: for the first time in Olympic history, the Athletes' Parade took place in a widespread format, simultaneously connecting Milan (San Siro), Livigno (Snowboard Park), Predazzo (Ski Jumping Stadium) and Cortina (city centre). The 92 delegations, starting with Greece and ending with Italy, paraded in the locations closest to their competition venues, transforming the Italian landscape into a single shared stage.

This was a symbolic choice by Italy, which, to represent the unity of the country, exceptionally fielded four flag bearers: two in Milan. Arianna Fontana and Federico Pellegrino and two in Cortina, Federica Brignone and Amos Mosaner.

3. Journey through time: from Munari's gestures to great history

The moment: an artistic segment that retraced 100 years of Winter Olympics, with a performance by Brenda Lodigiani dedicated to the language of Italian gestures as described by Bruno Munari.

Speeches and protocol

Giovanni Malagò (President of the Organising Committee):

'Tonight, Italy opens its arms to the world,' said Malagò, recalling previous Games hosted by our country. 'We are ready to write Olympic history once again.'

Kirsty Coventry (President of the International Olympic Committee):

'You've made it, be proud of getting this far. Now give it your best, enjoy yourselves, savour every moment. You will give us something special, you will make us dream, you will show us that strength is not just a matter of victories but also of courage, empathy and heart. You will teach us to get back up, no matter how hard the fall.'

Sergio Mattarella (President of the Italian Republic): from the VIP stand, the President of the Italian Republic, Sergio Mattarella, officially declared the Milan Cortina 2026 Winter Olympic Games open.

4. The Journey of the Olympic Flame: a collective relay race

The moment: the visual story of the Flame's journey, which started in Olympia and crossed the Italian landscape to unite Milan, Cortina and all the venues of the Games, was shown on the screens.

Inside the San Siro stadium, the "Essential" Torch was the star of a symbolic relay: first carried by three torchbearers, including Paola Egonu, then entrusted to two successive groups of three Italian athletes to leave the stadium accompanied by a performance by Andrea Bocelli.

5. The Olympic Flag: a bridge between generations and territories

The moment: the Olympic Flag made its solemn entrance following the widespread logic of the event. Carried by ten sporting legends and prominent personalities (eight in Milan and two in Cortina), the flag with the Five Rings passed through the symbolic locations of the Ceremony.

6. Lighting the Braziers

The moment: the last three torchbearers, Olympic alpine skiing champions Alberto Tomba and Deborah Compagnoni, simultaneously lit the Brazier at the Arco della Pace in Milan, while Olympic alpine skiing champion Sofia Goggia lit the Brazier in Piazza Angelo Dibona in Cortina.

Thus, the Olympic Flame united the urban dimension of Milan with the majesty of the Dolomites, a UNESCO World Heritage Site.

The Emotions of the Opening Ceremony of Milan Cortina 2026: Olympic Values in a Widespread Celebration (GOOGLE TRANSLATE)

Under the banner of Harmony, the XXV Olympic Winter Games connected Milan's San Siro Stadium in real time with the other regions integral to the great sporting event.

The Opening Ceremony of Milan Cortina 2026 marked a historic turning point: for the first time in the history of the Games, the Ceremony took place in a widespread format, involving Milan, Cortina, Livigno, Predazzo, and Antholz/Anterselva in a single, real-time orchestrated narrative.

The creative concept, entitled "Harmony," represented the ability to bring together different elements: city and mountain, man and nature.

To respect the principle of sustainability and minimize travel, the athletes were not gathered in Milan, but marched in the four venues closest to their respective Olympic Villages: Milan (San Siro Stadium), Cortina d'Ampezzo (Piazza Angelo Dibona), Livigno, and Predazzo.

During the evening, performances by artists, performers, and a wealth of highlights awaited: from Mariah Carey singing Domenico Modugno's "Nel Blu, dipinto di Blu," followed by one of her most iconic songs, "Nothing Is Impossible," to performances by Matilda De Angelis, Pierfrancesco Favino, and live performances led by Sabrina Impacciatore, Ghali, and Charlize Theron, the latter at the center of the canvas dedicated to the Olympic Truce and the reflection on the rejection of war.

Here are some of the moments to remember.

The evening's emotions

1. The National Anthem and the embrace of the Tricolor

The moment: the institutional protocol took place simultaneously in Milan and Cortina. The National Anthem was performed by Laura Pausini at the San Siro Stadium, accompanied by a captivating Mountain Choir in Cortina d'Ampezzo. Top model Vittoria Ceretti, who walked the runway alongside a group of models wearing creations designed by Giorgio Armani, was chosen to represent the city's creative and contemporary spirit. She carried the National Flag to the protocol stage, where it was entrusted to the Carabinieri Corps.

Simultaneously, in Cortina, a group of flag bearers, representing sporting tradition—Fulvio Valbusa, Giorgio Di Centa, Pietro Piller Cottrer, and Cristian Zorzi—performed the same gesture, handing over the Italian Flag to the Carabinieri Corps.

2. The Athletes' Parade: A Single Parade for the Regions

The Moment: For the first time in Olympic history, the Athletes' Parade took place across the Olympic circuit, simultaneously connecting Milan (San Siro), Livigno (Snowboard Park), Predazzo (Ski Jumping Stadium), and Cortina (City Center). The 92 delegations, starting in Greece and ending in Italy, marched at the sites closest to their respective competition venues, transforming the Italian geography into a single shared stage.

This was a symbolic choice by Italy, which, to represent the country's unity, exceptionally fielded four flag bearers: two in Milan, Arianna Fontana and Federico Pellegrino, and two in Cortina, Federica Brignone and Amos Mosaner.

3. Time Travel: From Munari's Gestures to Great History

The Moment: An artistic segment that retraced 100 years of the Winter Olympics, with a performance by Brenda Lodigiani dedicated to the language of Italian gestures as told by Bruno Munari.

Speeches and Protocol

Giovanni Malagò (President of the Organizing Committee):

"Tonight, Italy opens its arms to the world," said Malagò, recalling previous Games hosted by our country. "We are ready to write Olympic history again."

Kirsty Coventry (President of the International Olympic Committee):

"You made it, be proud to have come this far. Now give it your all, have fun, savor every moment. You will give us something special, you will make us dream, you will show us that strength is not just a question of victories but also of courage, empathy, and heart. You will teach us to get up again, no matter how hard the fall."

Sergio Mattarella (President of the Italian Republic): From the grandstand, the President of the Italian Republic, Sergio Mattarella, officially opened the Milan-Cortina 2026 Winter Olympic Games.

4. The Journey of the Olympic Flame: a collective relay

The moment: The screens displayed a visual narrative of the journey of the Flame, starting from Olympia and crossing the Italian landscapes to connect Milan, Cortina, and all the venues of the Games.

Inside the San Siro Stadium, the "Essential" Torch took center stage in a symbolic relay: first carried by three torchbearers, including Paola Egonu, then entrusted to two successive groups of three Italian athletes for the exit from the stadium accompanied by a performance by Andrea Bocelli.

5. The Olympic Flag: A Bridge Between Generations and Territories

The Moment: The Olympic Flag made its solemn entrance, following the event's widespread logic. Carried by ten sports legends and prominent figures (eight in Milan and two in Cortina), the banner with the Five Rings passed through the symbolic locations of the Ceremony.

6. The Lighting of the Braziers

The Moment: The final three torchbearers, Olympic Alpine skiing champions Alberto Tomba and Deborah Compagnoni, simultaneously lit the Brazier at the Arco della Pace in Milan, while Olympic Alpine skiing champion Sofia Goggia lit the Brazier in Piazza Angelo Dibona in Cortina.

Thus, the Olympic Flame united the urban dimension of Milan with the majesty of the Dolomites, a UNESCO World Heritage Site.

The Emotions of the Milan Cortina 2026 Opening Ceremony: Olympic Values in a Widespread Celebration (CHATGPT)

Under the banner of *Harmony*, the XXV Winter Olympic Games connected, in real time, Milan's San Siro Stadium with the other territories that were an integral part of the great sporting event.

The Milan Cortina 2026 Opening Ceremony marked a historic turning point: for the first time in the history of the Games, the Ceremony took shape as a widespread event, involving Milan, Cortina, Livigno, Predazzo, and Antholz/Anterselva in a single narrative orchestrated live.

The creative concept, titled *"Harmony"*, represented the ability to bring together different elements and make them coexist: city and mountains, humankind and nature.

To respect the principle of sustainability and minimize travel, the athletes were not gathered in Milan but instead paraded in the four venues closest to their respective Olympic Villages: Milan (San Siro Stadium), Cortina d'Ampezzo (Piazza Angelo Dibona), Livigno, and Predazzo.

Throughout the evening, performances by artists and performers delivered a series of standout moments: from Mariah Carey singing *Nel Blu, Dipinto di Blu* by Domenico Modugno, followed by one of her most iconic songs, *Nothing Is Impossible*, to performances by Matilda De Angelis and Pierfrancesco Favino, as well as live show moments led by Sabrina Impacciatore, Ghali, and Charlize Theron—the latter two at the center of the segment dedicated to the Olympic Truce and reflection on the rejection of war.

Here are some of the moments to remember.

The Emotions of the Evening

1. The National Anthem and the Embrace of the Tricolour

The moment: the institutional protocol took place simultaneously in Milan and Cortina. The National Anthem was performed by Laura Pausini at San Siro Stadium, accompanied by an evocative Mountain Choir in Cortina d'Ampezzo.
Top model Vittoria Ceretti, who walked alongside a group of models wearing creations designed by Giorgio Armani, was chosen to represent the city's creative and contemporary spirit. She carried the National Flag to the protocol stage, where it was entrusted to the Corazzieri Corps.

At the same time in Cortina, a group of flag bearers representing sporting tradition—Fulvio Valbusa, Giorgio Di Centa, Pietro Piller Cottrer, and Cristian Zorzi—performed the same gesture, handing the Italian Flag to the Carabinieri Corps.

2. The Athletes' Parade: One Parade for All Territories

The moment: for the first time in Olympic history, the Athletes' Parade took place in a widespread format, simultaneously linking Milan (San Siro), Livigno (Snowboard Park), Predazzo (Ski Jumping Stadium), and Cortina (city center).
The 92 delegations, starting with Greece and closing with Italy, paraded at the venues closest to their competition sites, turning Italy's geography into a single shared stage.

As a symbolic choice, Italy—representing national unity—exceptionally fielded four flag bearers: two in Milan, Arianna Fontana and Federico Pellegrino, and two in Cortina, Federica Brignone and Amos Mosaner.

3. A Journey Through Time: From Munari’s Gestures to Great History

The moment: an artistic segment retracing 100 years of Winter Olympics, featuring a performance by Brenda Lodigiani dedicated to Italian gesture language as narrated by Bruno Munari.

Speeches and Protocol

Giovanni Malagò (President of the Organizing Committee):
“Tonight Italy opens its arms to the world,” Malagò stated, recalling the previous Games hosted by the country. “We are ready to write Olympic history once again.”

Kirsty Coventry (President of the International Olympic Committee):
“You’ve made it—be proud of how far you’ve come. Now give your best, enjoy yourselves, savor every moment. You will give us something special, make us dream, and show us that strength is not only about victories, but also about courage, empathy, and heart. You will teach us how to get back up, no matter how hard the fall.”

Sergio Mattarella (President of the Italian Republic):
From the honorary stand, the President of the Italian Republic, Sergio Mattarella, officially declared the Milan Cortina 2026 Winter Olympic Games open.

4. The Journey of the Olympic Flame: A Collective Relay

The moment: the screens displayed the visual story of the Flame’s journey, which began in Olympia and traveled across Italy’s landscapes, ultimately uniting Milan, Cortina, and all the Games’ venues.

Inside San Siro Stadium, the *“Essential”* Torch took center stage in a symbolic relay: first carried by three torchbearers, including Paola Egonu, then passed to two subsequent groups of three Italian athletes for their exit from the stadium, accompanied by a performance by Andrea Bocelli.

5. The Olympic Flag: A Bridge Between Generations and Territories

The moment: the Olympic Flag made its solemn entrance following the event’s widespread logic. Carried by ten sports legends and prominent figures (eight in Milan and two in Cortina), the flag with the Five Rings crossed the Ceremony’s symbolic locations.

6. Lighting the Cauldrons

The moment: the final three torchbearers—Olympic alpine skiing champions Alberto Tomba and Deborah Compagnoni—simultaneously lit the cauldron at the Arco della Pace in Milan, while Olympic alpine skiing champion Sofia Goggia lit the cauldron in Piazza Angelo Dibona in Cortina.

In this way, the Olympic Flame united Milan’s urban dimension with the majesty of the UNESCO-listed Dolomites.

APPENDIX II

Table 1A – Analysed Translations (1956 article)

source	reference	deepl	google	chatgpt
I Giochi Olimpici si sono iniziati ufficialmente questa mattina alle 10.10, quando il Presidente della Repubblica nello Stadio del Ghiaccio, popolato di concorrenti e gremito di folla, si è alzato per pronunciare la formula di prammatica: «Dichiaro aperti i VII Giochi Olimpici Invernali di Cortina d'Ampezzo celebranti la XVI Olimpiade dell'era moderna».	The Winter Olympic Games officially began this morning at 10:10, when the President of the Republic stood up in the Ice Stadium, filled with athletes and spectators, to pronounce the customary oath: "I declare open the VII Winter Olympic Games of Cortina d'Ampezzo, celebrating the XVI Olympiad of the modern era".	The Olympic Games officially began this morning at 10.10 a.m., when the President of the Republic stood up in the Ice Stadium, filled with competitors and packed with crowds, to pronounce the customary formula: "I declare open the VII Winter Olympic Games in Cortina d'Ampezzo, celebrating the XVI Olympiad of the modern era".	The Olympic Games officially began this morning at 10:10, when the President of the Republic stood in the packed Ice Stadium to pronounce the formal oath: "I declare open the VII Olympic Winter Games of Cortina d'Ampezzo, celebrating the XVI Olympiad of the modern era."	The Olympic Games officially began this morning at 10:10, when the President of the Republic, in the Ice Stadium filled with competitors and packed with spectators, rose to pronounce the customary formula: "I declare open the VII Winter Olympic Games of Cortina d'Ampezzo, celebrating the XVI Olympiad of the modern era."
Poi la bandiera dai cerchi di cinque colori è salita lentamente sul pennone principale e l'artiglieria ha sparato tre salve di cannone che a lungo si sono ripercosse tra le gole dei monti incombenti.	Then the flag with the five-coloured rings slowly rose on the main flagpole and the artillery fired three cannon salvoes that echoed for a long time through the gorges of the sorrounding mountains.	Then the flag with the five coloured circles slowly rose on the main flagpole and the artillery fired three cannon salutes that echoed for a long time in the gorges of the looming mountains.	Then the five-colored circled flag slowly rose to the main mast and the artillery fired three cannon salutes that echoed for a long time through the gorges of the looming mountains.	Then the flag with the five-colored rings slowly rose on the main flagpole, and the artillery fired three cannon salvos that echoed for a long time among the surrounding mountain gorges.
Non è stato difficile incolonnarle: in testa la Grecia, perché questo onore spetta al popolo che fondò le Olimpiadi, poi tutte le altre in ordine alfabetico fino all'Italia, sistemata in coda per il fatto di essere padrona di casa.	It was not difficult to line them up: Greece led the way, because this honor belongs to the people who founded the Olympics, followed by all the other nations in alphabetical order, with Italy placed at the end of the queue because it is the host country.	It was not difficult to line them up: Greece at the head, because this honour belongs to the people who founded the Olympics, then all the others in alphabetical order up to Italy, placed at the end because it was the host country.	It wasn't difficult to line them up: Greece led the way, because this honor belongs to the people who founded the Olympics, followed by all the others in alphabetical order, ending with Italy, placed at the end of the queue because it was the host country.	It was not difficult to line them up: Greece at the head, since this honor belongs to the people who founded the Olympics, followed by all the others in alphabetical order, with Italy at the rear as the host nation.
Questi ultimi sembrano troppi; facciamo notare che sono invece pochi rispetto alla reale quantità degli «ufficiali» scesi a Cortina al seguito dei concorrenti.	The latter seem too many; we note, however, that they are actually few compared to the real number of team officials who have come to Cortina to follow the athletes.	The latter seem too many; we would point out that they are actually few compared to the real number of "officials" who came to Cortina accompanying the competitors.	The latter number may seem excessive; it should be noted that they are actually few compared to the actual number of "officials" who descended on Cortina to accompany the competitors.	The latter seem too many; we note, however, that they are actually few compared with the real number of "officials" who have come to Cortina in the athletes' wake.
Qualche esempio preso a caso: la Corea del Sud, che evidentemente non bada a spese, ha mandato 4 atleti e 3 accompagnatori; il Giappone rispettivamente 10 e 10 (si vede che l'Italia attira enormemente gli orientali).	A few random examples: South Korea, which clearly spared no expense, sent four athletes and three team officials; Japan sent ten and ten respectively (it is evident that Italy is a huge attraction for Asians).	A few random examples: South Korea, which clearly spared no expense, sent four athletes and three accompanying persons; Japan sent ten and ten respectively (it is clear that Italy is a huge attraction for Asians);	A few random examples: South Korea, evidently sparing no expense, sent four athletes and three accompanying persons; Japan 10 and 10 respectively (it's clear that Italy is a huge draw for Easterners);	A few random examples: South Korea, which clearly spares no expense, sent 4 athletes and 3 officials; Japan, respectively 10 and 10 (Italy evidently exerts enormous attraction on Easterners);
Di fronte a queste cifre il Comitato organizzatore si è preoccupato ed ha deciso misure drastiche;	Faced with these numbers, the organising committee became concerned and decided on drastic measures.	Faced with these figures, the organising committee became concerned and decided on drastic measures;	Faced with these numbers, the organizing committee became concerned and decided on drastic measures;	Faced with these figures, the Organizing Committee grew concerned and decided on drastic measures:
Molti accompagnatori sono quindi rimasti in tribuna o in albergo, ma lo spettacolo ci ha guadagnato in snellezza ed eleganza.	Many of their team officials therefore remained in the stands or at their hotels, but the show was more streamlined and elegant.	Many accompanying persons therefore remained in the stands or in their hotels, but the show gained in streamlined elegance.	Many of their companions therefore remained in the stands or at their hotels, but the spectacle was enhanced in both elegance and speed.	Many officials therefore remained in the stands or at their hotels, but the spectacle gained in slimness and elegance.
Così la colonna variopinta, preceduta da un gruppo di 25 alfieri (allievi dell'Istituto superiore di educazione fisica di Roma) reggenti le bandiere olimpiche, è sfilata per le vie di Cortina ed ha raggiunto il piazzale esterno dello Stadio del Ghiaccio.	Thus, the colorful parade marched through the streets of Cortina and reached the square outside the Ice Stadium, preceded by a group of 25 flagbearers (students from the Higher Institute of Physical Education in Rome) holding the Olympic flags.	Thus, the colourful column, preceded by a group of 25 flag bearers (students from the Higher Institute of Physical Education in Rome) carrying the Olympic flags, paraded through the streets of Cortina and reached the square outside the Ice Stadium.	Thus, the colorful convoy, preceded by a group of 25 standard-bearers (students from the Higher Institute of Physical Education in Rome) holding Olympic flags, paraded through the streets of Cortina and reached the square outside the Ice Stadium.	Thus the colorful column, preceded by a group of 25 standard-bearers (students of the Higher Institute of Physical Education of Rome) carrying the Olympic flags, paraded through the streets of Cortina and reached the open area in front of the Ice Stadium.
Verso il termine tre squadroni: i russi, gli americani, gli italiani, che sono i gruppi più numerosi. All'apparire dei nostri l'urlo della folla diviene tempesta: non sappiamo se gli amplificatori degli apparecchi televisivi hanno potuto registrarlo con esattezza. Guida i nostri il saltatore Nilo Zandanel.	Toward the end we see the three largest delegations: the Russians, the Americans, and the Italians. When our team appeared, the crowd erupted in a thunderous roar; we do not know whether television amplifiers were able to record it accurately. Leading our athletes is the ski jumper Nilo Zandanel.	Towards the end, there are three large teams: the Russians, the Americans and the Italians, which are the largest groups. When our team appeared, the crowd's roar became a storm: we do not know if the television amplifiers were able to record it accurately. Our team was led by the jumper Nilo Zandanel.	Towards the end, three teams emerge: the Russians, the Americans, and the Italians, who are the largest groups. When our team appears, the roar of the crowd escalates into a storm: we don't know if the television amplifiers were able to accurately record it. Leading our team is high jumper Nilo Zandanel.	Toward the end come three large squads: the Russians, the Americans, and the Italians, the largest groups of all. At the appearance of our own team, the roar of the crowd becomes a tempest; we do not know whether television amplifiers were able to record it accurately. Leading our athletes is the ski jumper Nilo Zandanel.
Nella disavventura ha la presenza di spirito di tenere alta la fiaccola, e questa pertanto non si spegne.	In his misfortune, he has the presence of mind to hold the torch high, and thus the flame is kept alive.	In his misfortune, he has the presence of mind to hold the torch high, and so it does not go out.	In his misfortune, he had the presence of mind to hold the torch high, and so it remained.	In his mishap he has the presence of mind to keep the torch raised, and thus it does not go out.
Ora una figurina lo segue sul palco: la nostra Giuliana Minuzzo, medaglia di bronzo ad Oslo, campionessa di discesa.	Now another athlete follows him on stage: our very own Giuliana Minuzzo, bronze medallist in Oslo, downhill champion.	Now a figure follows him onto the stage: our Giuliana Minuzzo, bronze medallist in Oslo, downhill champion.	Now a figurine follows him onto the stage: our own Giuliana Minuzzo, bronze medalist in Oslo and downhill champion.	Now a small figure follows him onto the platform: our Giuliana Minuzzo, bronze medalist at Oslo, downhill champion.
Forse è più emozionata che alla partenza di una gara di grande impegno. Afferra con la destra un lembo della bandiera italiana e mentre gli altri vessilli si piegano sino a sfiorare la pista, pronuncia la formula del rito	Perhaps she is more emotional than at the start of a highly demanding race. With her right hand she grasps a fold of the Italian flag, and while the other flags bow until they nearly touch the track, she pronounces the oath:	Perhaps she is more excited than at the start of a major competition. She grabs a corner of the Italian flag with her right hand and, while the other flags are lowered to touch the track, she recites the ritual formula:	Perhaps she's more excited than she is at the start of a challenging race. He grasps a corner of the Italian flag in his right hand, and as the other flags fold until they touch the track, he pronounces the ritual formula:	Perhaps she is more emotional than at the start of a highly demanding race. With her right hand she grasps a fold of the Italian flag, and while the other banners bow until they nearly touch the ice, she pronounces the ritual formula:
Da questo momento la migliore gioventù del mondo si batte pacificamente per un primato di gloria sportiva.	From this moment on, the finest youth of the world will compete peacefully to gain a title of sporting glory.	From this moment on, the best young people in the world will compete peacefully for sporting glory.	From this moment, the finest youth in the world will fight peacefully for a record of sporting glory.	From this moment, the finest youth of the world will compete peacefully for a title of sporting glory.

Table 2A – Metrics Results

Sentence Number	MT Systems	COMET	TER
1	DeepL	0,73	0,18
1	Google	0,70	0,25
1	ChatGPT	0,72	0,27
2	DeepL	0,68	0,18
2	Google	0,61	0,27
2	ChatGPT	0,82	0,27
3	DeepL	0,73	0,32
3	Google	0,82	0,15
3	ChatGPT	0,74	0,30
4	DeepL	0,71	0,37
4	Google	0,65	0,53
4	ChatGPT	0,75	0,20
5	DeepL	0,83	0,11
5	Google	0,73	0,37
5	ChatGPT	0,65	0,54
6	DeepL	0,97	0,14
6	Google	0,98	0,14
6	ChatGPT	0,92	0,29
7	DeepL	0,48	0,43
7	Google	0,27	0,39
7	ChatGPT	0,35	0,39
8	DeepL	0,65	0,36
8	Google	0,67	0,26
8	ChatGPT	0,40	0,54
9	DeepL	0,79	0,53
9	Google	0,78	0,59
9	ChatGPT	0,80	0,47
10	DeepL	0,52	0,29
10	Google	0,68	0,33
10	ChatGPT	0,22	0,38
11	DeepL	0,85	0,33
11	Google	0,83	0,44
11	ChatGPT	0,79	0,56
12	DeepL	0,77	0,38
12	Google	0,75	0,50
12	ChatGPT	0,79	0,10
13	DeepL	0,77	0,45
13	Google	0,77	0,35
13	ChatGPT	0,83	0,20

Table 1B – Analysed Translations (2026 article)

source	reference	deepl	google	chatgpt
Sotto il segno dell'Armonia, i XXV Giochi Olimpici Invernali hanno connesso in tempo reale lo Stadio San Siro di Milano con gli altri territori parte integrante del grande evento sportivo.	Sotto il segno dell'Armonia, i XXV Giochi Olimpici Invernali hanno connesso in tempo reale lo Stadio San Siro di Milano con gli altri territori parte integrante del grande evento sportivo.	Sotto il segno dell'Armonia, i XXV Giochi Olimpici Invernali hanno connesso in tempo reale lo Stadio San Siro di Milano con gli altri territori parte integrante del grande evento sportivo.	Sotto il segno dell'Armonia, i XXV Giochi Olimpici Invernali hanno connesso in tempo reale lo Stadio San Siro di Milano con gli altri territori parte integrante del grande evento sportivo.	Under the banner of *Harmony*, the XXV Winter Olympic Games connected, in real time, Milan's San Siro Stadium with the other territories that were an integral part of the great sporting event.
Sotto il segno dell'Armonia, i XXV Giochi Olimpici Invernali hanno connesso in tempo reale lo Stadio San Siro di Milano con gli altri territori parte integrante del grande evento sportivo.	Sotto il segno dell'Armonia, i XXV Giochi Olimpici Invernali hanno connesso in tempo reale lo Stadio San Siro di Milano con gli altri territori parte integrante del grande evento sportivo.	Sotto il segno dell'Armonia, i XXV Giochi Olimpici Invernali hanno connesso in tempo reale lo Stadio San Siro di Milano con gli altri territori parte integrante del grande evento sportivo.	Sotto il segno dell'Armonia, i XXV Giochi Olimpici Invernali hanno connesso in tempo reale lo Stadio San Siro di Milano con gli altri territori parte integrante del grande evento sportivo.	To respect the principle of sustainability and minimize travel, the athletes were not gathered in Milan but instead paraded in the four venues closest to their respective Olympic Villages: Milan (San Siro Stadium), Cortina d'Ampezzo (Piazza Angelo Dibona), Livigno, and Predazzo.
Sotto il segno dell'Armonia, i XXV Giochi Olimpici Invernali hanno connesso in tempo reale lo Stadio San Siro di Milano con gli altri territori parte integrante del grande evento sportivo.	Sotto il segno dell'Armonia, i XXV Giochi Olimpici Invernali hanno connesso in tempo reale lo Stadio San Siro di Milano con gli altri territori parte integrante del grande evento sportivo.	Sotto il segno dell'Armonia, i XXV Giochi Olimpici Invernali hanno connesso in tempo reale lo Stadio San Siro di Milano con gli altri territori parte integrante del grande evento sportivo.	Sotto il segno dell'Armonia, i XXV Giochi Olimpici Invernali hanno connesso in tempo reale lo Stadio San Siro di Milano con gli altri territori parte integrante del grande evento sportivo.	Top model Vittoria Ceretti, who walked alongside a group of models wearing creations designed by Giorgio Armani, was chosen to represent the city's creative and contemporary spirit. She carried the National Flag to the protocol stage, where it was entrusted to the Corazzieri Corps.
Sotto il segno dell'Armonia, i XXV Giochi Olimpici Invernali hanno connesso in tempo reale lo Stadio San Siro di Milano con gli altri territori parte integrante del grande evento sportivo.	Sotto il segno dell'Armonia, i XXV Giochi Olimpici Invernali hanno connesso in tempo reale lo Stadio San Siro di Milano con gli altri territori parte integrante del grande evento sportivo.	Sotto il segno dell'Armonia, i XXV Giochi Olimpici Invernali hanno connesso in tempo reale lo Stadio San Siro di Milano con gli altri territori parte integrante del grande evento sportivo.	Sotto il segno dell'Armonia, i XXV Giochi Olimpici Invernali hanno connesso in tempo reale lo Stadio San Siro di Milano con gli altri territori parte integrante del grande evento sportivo.	The 92 delegations, starting with Greece and closing with Italy, paraded at the venues closest to their competition sites, turning Italy's geography into a single shared stage.
Sotto il segno dell'Armonia, i XXV Giochi Olimpici Invernali hanno connesso in tempo reale lo Stadio San Siro di Milano con gli altri territori parte integrante del grande evento sportivo.	Sotto il segno dell'Armonia, i XXV Giochi Olimpici Invernali hanno connesso in tempo reale lo Stadio San Siro di Milano con gli altri territori parte integrante del grande evento sportivo.	Sotto il segno dell'Armonia, i XXV Giochi Olimpici Invernali hanno connesso in tempo reale lo Stadio San Siro di Milano con gli altri territori parte integrante del grande evento sportivo.	Sotto il segno dell'Armonia, i XXV Giochi Olimpici Invernali hanno connesso in tempo reale lo Stadio San Siro di Milano con gli altri territori parte integrante del grande evento sportivo.	As a symbolic choice, Italy—representing national unity—exceptionally fielded four flag bearers: two in Milan, Arianna Fontana and Federico Pellegrino, and two in Cortina, Federica Brignone and Amos Mosaner.
Sotto il segno dell'Armonia, i XXV Giochi Olimpici Invernali hanno connesso in tempo reale lo Stadio San Siro di Milano con gli altri territori parte integrante del grande evento sportivo.	Sotto il segno dell'Armonia, i XXV Giochi Olimpici Invernali hanno connesso in tempo reale lo Stadio San Siro di Milano con gli altri territori parte integrante del grande evento sportivo.	Sotto il segno dell'Armonia, i XXV Giochi Olimpici Invernali hanno connesso in tempo reale lo Stadio San Siro di Milano con gli altri territori parte integrante del grande evento sportivo.	Sotto il segno dell'Armonia, i XXV Giochi Olimpici Invernali hanno connesso in tempo reale lo Stadio San Siro di Milano con gli altri territori parte integrante del grande evento sportivo.	An artistic segment retracing 100 years of Winter Olympics, featuring a performance by Brenda Lodigiani dedicated to Italian gesture language as narrated by Bruno Munari.
Sotto il segno dell'Armonia, i XXV Giochi Olimpici Invernali hanno connesso in tempo reale lo Stadio San Siro di Milano con gli altri territori parte integrante del grande evento sportivo.	Sotto il segno dell'Armonia, i XXV Giochi Olimpici Invernali hanno connesso in tempo reale lo Stadio San Siro di Milano con gli altri territori parte integrante del grande evento sportivo.	Sotto il segno dell'Armonia, i XXV Giochi Olimpici Invernali hanno connesso in tempo reale lo Stadio San Siro di Milano con gli altri territori parte integrante del grande evento sportivo.	Sotto il segno dell'Armonia, i XXV Giochi Olimpici Invernali hanno connesso in tempo reale lo Stadio San Siro di Milano con gli altri territori parte integrante del grande evento sportivo.	Sergio Mattarella (President of the Italian Republic): from the honorary stand, the President of the Italian Republic, Sergio Mattarella, officially declared the Milan Cortina 2026 Winter Olympic Games open.
Sotto il segno dell'Armonia, i XXV Giochi Olimpici Invernali hanno connesso in tempo reale lo Stadio San Siro di Milano con gli altri territori parte integrante del grande evento sportivo.	Sotto il segno dell'Armonia, i XXV Giochi Olimpici Invernali hanno connesso in tempo reale lo Stadio San Siro di Milano con gli altri territori parte integrante del grande evento sportivo.	Sotto il segno dell'Armonia, i XXV Giochi Olimpici Invernali hanno connesso in tempo reale lo Stadio San Siro di Milano con gli altri territori parte integrante del grande evento sportivo.	Sotto il segno dell'Armonia, i XXV Giochi Olimpici Invernali hanno connesso in tempo reale lo Stadio San Siro di Milano con gli altri territori parte integrante del grande evento sportivo.	Inside San Siro Stadium, the *"Essential"* Torch took center stage in a symbolic relay: first carried by three torchbearers, including Paola Egonu, then passed to two subsequent groups of three Italian athletes for their exit from the stadium, accompanied by a performance by Andrea Bocelli.
Sotto il segno dell'Armonia, i XXV Giochi Olimpici Invernali hanno connesso in tempo reale lo Stadio San Siro di Milano con gli altri territori parte integrante del grande evento sportivo.	Sotto il segno dell'Armonia, i XXV Giochi Olimpici Invernali hanno connesso in tempo reale lo Stadio San Siro di Milano con gli altri territori parte integrante del grande evento sportivo.	Sotto il segno dell'Armonia, i XXV Giochi Olimpici Invernali hanno connesso in tempo reale lo Stadio San Siro di Milano con gli altri territori parte integrante del grande evento sportivo.	Sotto il segno dell'Armonia, i XXV Giochi Olimpici Invernali hanno connesso in tempo reale lo Stadio San Siro di Milano con gli altri territori parte integrante del grande evento sportivo.	The final three torchbearers—Olympic alpine skiing champions Alberto Tomba and Deborah Compagnoni—simultaneously lit the cauldron at the Arco della Pace in Milan, while Olympic alpine skiing champion Sofia Goggia lit the cauldron in Piazza Angelo Dibona in Cortina.
Sotto il segno dell'Armonia, i XXV Giochi Olimpici Invernali hanno connesso in tempo reale lo Stadio San Siro di Milano con gli altri territori parte integrante del grande evento sportivo.	Sotto il segno dell'Armonia, i XXV Giochi Olimpici Invernali hanno connesso in tempo reale lo Stadio San Siro di Milano con gli altri territori parte integrante del grande evento sportivo.	Sotto il segno dell'Armonia, i XXV Giochi Olimpici Invernali hanno connesso in tempo reale lo Stadio San Siro di Milano con gli altri territori parte integrante del grande evento sportivo.	Sotto il segno dell'Armonia, i XXV Giochi Olimpici Invernali hanno connesso in tempo reale lo Stadio San Siro di Milano con gli altri territori parte integrante del grande evento sportivo.	In this way, the Olympic Flame united Milan's urban dimension with the majesty of the UNESCO-listed Dolomites.

Table 2B – Metrics Results

Sentence Number	MT Systems	COMET	TER
1	DeepL	0,85	0,30
1	Google_translate	0,82	0,37
1	ChatGPT	0,83	0,27
2	DeepL	0,84	0,15
2	Google_translate	0,81	0,26
2	ChatGPT	0,83	0,21
3	DeepL	0,84	0,15
3	Google_translate	0,84	0,23
3	ChatGPT	0,84	0,25
4	DeepL	0,83	0,14
4	Google_translate	0,83	0,25
4	ChatGPT	0,79	0,32
5	DeepL	0,69	0,53
5	Google_translate	0,70	0,40
5	ChatGPT	0,73	0,37
6	DeepL	0,74	0,15
6	Google_translate	0,72	0,18
6	ChatGPT	0,67	0,39
7	DeepL	0,77	0,15
7	Google_translate	0,73	0,15
7	ChatGPT	0,78	0,15
8	DeepL	0,69	0,33
8	Google_translate	0,73	0,17
8	ChatGPT	0,77	0,09
9	DeepL	0,76	0,07
9	Google_translate	0,69	0,10
9	ChatGPT	0,82	0,15
10	DeepL	0,94	0,29
10	Google_translate	0,94	0,29
10	ChatGPT	0,88	0,29