Lexical Variation in English–Italian News Translation: A Comparative Study of Google Translate and ChatGPT-5.2

Aurora Trapella[1,2] Lieve Macken[2] and Alessandra Molino[1]
[1]Department of Foreign Languages, Literatures, and Modern Cultures, University of Turin, Italy
[2]Language and Translation Technology Team (LT[3]), Department of Translation, Interpreting and Communication, Ghent University, Belgium

[1]aurora.trapella@unito.it

[2]lieve.macken@ugent.be

[3]alessandra.molino@unito.it

Abstract

This study investigates stylistic features in English–Italian machine translation (MT) by comparing the outputs of Google Translate (GT) and ChatGPT-5.2 (GPT). The analysis adopts a quantitative corpus-based approach to studying news reports and opinion articles for which parallel translations produced by both systems were collected via API. In particular, the focus is on the lexical variety of the translation equivalents of three high-frequency mental verbs: *know*, *think*, and *believe*. Concordance lines were analyzed and annotated according to the equivalents used, enabling a systematic comparison of lexical range across systems and the two news genres. The results show that the two systems seem to adopt similar translation solutions, but they differ in the quantitative distribution of translation equivalents. While GT exhibits greater sensitivity to morphological variation, GPT tends to rely on a more restricted set of equivalents and a wider range of phrasal and clausal reformulations. These findings contribute to studies on machine translationese (MTese) for the English–Italian language pair, with implications for the study of variation in MT and Artificial Intelligence (AI) literacy.

1 Introduction

This study investigates the stylistic features of automatic translations of a corpus of news texts comprising the two subgenres of news reports and opinion articles. The analysis focuses on news translated from English into Italian using Google Translate – a Neural MT (NMT) tool – and ChatGPT-5.2 – a general-purpose Large Language Model (LLM). Studies on machine-translated language often refer to it as an "artificially impoverished" language (Vanmassenhove et al., 2021), also named machine translationese (MTese) (Daems et al., 2021; De Clercq et al., 2021), which has become a widely investigated topic in disciplines such as Computational Linguistics, Natural Language Processing (NLP), and Translation Studies. Despite extensive research on the features of statistical and neural MT, the language output produced by LLMs has only recently begun to attract scholarly attention (Kong and Macken 2025b). However, to the best of our knowledge, little to no research has been carried out on the English–Italian language combination, although recently the WMT Shared Task (Kocmi et al., 2025) reported results indicating that news translated from English into Italian by the top-performing systems still ranked in the lowest positions. This underscores the necessity for additional research to be conducted on the characteristics of machine-translated Italian, with a view to enhancing the performance of training models and

supporting the training of translators and post-editors to recognize machine interference and preserve idiomatic language.

Such investigations are also required in light of the increasing use of AI tools in journalistic translation workflows and other non-professional settings. With the increasing integration of AI tools into different phases of the journalistic value chain (Cools and Diakopoulos, 2024), MT tools have become fundamental to journalistic workflows, and more recently, Generative AI (GenAI) tools, especially chatbots like OpenAI's ChatGPT, have been incorporated into journalistic practices for translation purposes (Roush 2023), with reference as well to the Italian context (Cagnan, 2025). While some news outlets rely on professional post-editing (Bullard 2023) or on in-house editing and journalist review ('journalist-translator') (Cagnan, 2025), others rely on fully automated translation (Roush, 2023) and on automatic content generation through transcription and automatic speech recognition (ASR) (The New York Times Trust Team 2024). However, there seems to be a lack of a theoretical framework due to the opportunistic use of AI technologies on a case-by-case basis (Caswell, 2019), together with a difficulty in acquiring AI literacy (Murru and Carlo, 2024).

Without the development of AI literacy (Bowker 2024), over-reliance on these tools by journalists and editors might result in information drift and the unintentional introduction of bias and stance-altered expressions, while lay users who use NMT tools or plugins (e.g., Google Translate) and LLMs (e.g., ChatGPT) to translate online news texts for personal use might encounter a less nuanced and potentially standardized language (Hermand, 2014).

Moreover, news discourse is characterized by frequent references to cognitive and epistemic states, as reported information is often mediated through evaluation and attribution. In this context, cognitive verbs as defined by Biber et al. (1999) play a key role in signaling degrees of certainty and perspective. These verbs may pose challenges for MT systems, which can struggle to capture subtle differences in epistemic force and contextual usage. For this reason, mental cognitive verbs are selected as the focus of the present study.

This study is an exploratory investigation within a larger project examining the linguistic and stylistic features of Italian MTese. The project has two main objectives: (1) to compare raw, non-post-edited translations generated by NMT systems and LLMs in order to identify whether and how they differ in their rendering of English source texts (ST), considering that NMT is a translation-specific technology whereas LLMs were not originally developed for translation tasks; and (2) to compare the output produced by NMT and LLMs with a comparable corpus of Italian news texts written before the advent of AI. The overall aim of the project is to provide an empirical description of the features of Italian MTese and to identify possible signs of standardisation in machine-translated output (Hermand, 2014). The findings will have implications for translator and post-editor training, as well as for AI literacy development among both professionals (e.g. journalists and editors) and general users.

The present study aims to answer the following research questions:

(1) How do Google Translate (GT) and ChatGPT (GPT) differ in lexical variation when translating English mental cognitive verbs (*know*, *believe*, *think*) into Italian?

(2) Does genre play a role in the extent to which these stylistic tendencies differ?

2 Literature review

Style has long been recognized as a central dimension of linguistic expression, commonly defined as a writer's individual mode of expression. As Murtaza and Qasmi (2013) argue, style encompasses a wide range of choices operating along both paradigmatic and syntagmatic axes, including the choice of lexical items, the use of figurative language, syntactic and phrasal structuring, and the organization of discourse in paragraphs. From this perspective, stylistic variation emerges from systematic patterns of linguistic choice, which can be observed and measured across texts.

In the context of MT, several works point to the frequent reduction of lexical and stylistic variation by neural MT tools. Indeed, machine-translated language is often characterized by processes of standardization also referred to as 'lexical convergence' (Hermand, 2014), whereby target language (TL) output tends to use a simplified range of vocabulary and to conform to more neutral and conventional patterns. This tendency is closely related to findings that MT systems frequently reduce lexical and stylistic diversity. For instance, Vanmassenhove et al. (2019) show that NMT systems tend to generalize and avoid colloquialisms,

leading to a loss of nuance, while also exhibiting biases that favor more frequent lexical items at the expense of less common ones. This "algorithmic bias" contributes to a narrowing of the lexical repertoire available in translated output.

Further evidence of reduced diversity in MT comes from studies comparing translated texts with original language data. Translations produced by different MT architectures have been shown to be consistently less lexically diverse than their training data (Vanmassenhove et al., 2021) and to contain fewer low-frequency items such as hapax legomena, alongside occasional non-standard or non-sense forms (De Clercq et al., 2021). This reduction in lexical variation has been described as a form of lexical impoverishment, whereby the range of available equivalents is simplified (Espinosa-Giménez and Forcada, 2024). Similarly, Loock (2020) and Jiang and Niu (2022) observe that MT output tends to over-represent certain linguistic features and lexical items, often due to source language interference or a lack of sensitivity to TL norms. Moreover, as Webster et al. (2020) note, MT systems may produce translations that are semantically comprehensible but stylistically inadequate, failing to capture appropriate register and more idiomatic formulations.

At the same time, genre has been shown to play a significant role in lexical variation (Niu and Jiang, 2024). Moreover, recent research has begun to explore differences between NMT and LLMs. While NMT has been associated with tendencies toward standardization and reduced variation, LLM-based systems appear to exhibit greater stylistic flexibility. Kong and Macken (2025a) observe that LLM outputs contain a higher frequency of descriptive lexical items, which are generally linked to increased textual vividness and specificity. In a subsequent study, Kong and Macken (2025b) report higher levels of lexical diversity in LLM outputs compared to NMT, suggesting a broader and more varied lexical repertoire. Despite these advances, capturing stylistic nuance remains a challenge for MT systems, which continue to struggle in modelling context-sensitive stylistic choices (Daems, 2022). The majority of these studies focus on language combinations such as English–French, English–Dutch, or Chinese–English, while the English–Italian pair remains underexplored in this respect. Against this background, the present study contributes to the growing body of research on stylistic variation in MT by comparing how different systems realize lexical variation across news genres.

3 Methodology

3.1 Corpus description

The English STs corpus contains approximately 340,000 words of published news retrieved from two pre-existing resources: the Dutch Parallel Corpus (DPC) (Macken et al., 2010) and the EMA ("editorial material") corpus (Molino, 2010), which were compiled for previous research projects. The texts contained in these two resources were published in similar time periods: DPC collects texts published between 2003 and 2008, while EMA contains texts published between 2002 and 2004. As the texts were published before AI-assisted writing and translation became widespread, their selection avoids potential contamination from machine-generated language. To ensure stylistic and register consistency, only news published by British broadsheet newspapers present in these two corpora was selected for the study. Table 1 provides a description of the ST corpus.

As the texts selected for the English STs corpus were encoded differently, they were converted to UTF-8 using a Python script, and a formatting function was added to insert missing spaces after full stops, improving readability and facilitating the subsequent analysis. Further preliminary processing produced clean plain-text versions, removing metadata from the DPC and manually cleaning the EMA corpus of headlines, dates, and author information.

The corpus was pre-processed through Python using the spaCy (Honnibal et al., 2023) NLP framework. The script implemented language-specific models (en_core_web_lg for the English STs and it_core_news_lg for the Italian translations) to perform tokenization, lemmatization, Part-of-Speech (POS) tagging and Named Entity Recognition (NER), while the dependency parser and sentence segmenter were disabled, as the input data was already pre-segmented into a one-sentence-per-line format. The data was then serialized into a structured JSON format to facilitate subsequent quantitative and comparative analysis.

	DPC (n. words)	EMA (n. words)	Overall ST corpus (n. words)
News reports	182,622	0	182,622
The Guardian	71,337	0	71,337
The Independent	111,285	0	111,285
Opinion articles	77,057	79,370	156,427
Financial Times	0	2,544	2,544
The Daily Telegraph	0	18,072	18,072
The Guardian	17,516	31,608	49,124
The Independent	59,541	14,922	74,463
The Spectator	0	12,224	12,224
Total n. of words	259,679	79,370	**339,049**

Table 1. ST corpus description: total number of words across news genres (news reports and opinion articles) and each of the original corpora (DPC and EMA).

The TT corpora to be examined within the quantitative analysis of concordance lines comprise translations performed automatically via Python scripts for the free version of Google Translate API and the licensed ChatGPT-5.2 API in March 2026. Python scripts included a two-second pause between requests and a retry mechanism of up to three times. These functions prevent overloading the systems and reduce the risk of incomplete translations. For large files, the scripts for both systems implement a paragraph-based chunking strategy to preserve contextual coherence during processing. The following prompt was given to perform the translation using the licensed ChatGPT-5.2 API: "You are a professional translator. Translate the following English text to Italian. Preserve the formatting, line breaks, and structure exactly. Only provide the translation, no explanations or additional text."

3.2 Research design

This study adopts a quantitative corpus-based approach to analyze lexical and stylistic variation in MT output. The analysis is based on concordance lines extracted for a set of selected ST verbs and focuses on the distribution of translation equivalents. Specifically, it assesses the extent to which the two systems differ in their lexical variability, operationalized as the number of distinct translation equivalents associated with each verb, and what differences can be observed between the MT systems and the news genres. In addition, differences in the structural realization of translation choices across the two systems and genres are discussed.

As this research serves as an exploratory study to assess whether a quantitative distribution of translation equivalents can effectively distinguish between the stylistic features of tools with different architectures, we focus on a small set of verbs to ensure accurate manual analysis. Such methodology will then be applied to a broader semantic range of verbs and translations produced by other tools in subsequent research phases.

3.3 Selection of linguistic variables

Mental verbs were selected as the focus of analysis based on their distribution across written registers in the *Longman Grammar of Spoken and Written English* (Biber et al., 1999). In their seminal work, Biber et al. (1999) argue that verbs can be classified according to their semantic domain, identifying seven categories: activity verbs, communication verbs, mental verbs, causative verbs, verbs of simple occurrence, verbs of existence or relationship and aspectual verbs. Overall, activity verbs constitute the largest proportion of verbs among the domains of conversation, fiction, news and academic prose, but their primarily descriptive function makes them less informative for a stylistic comparison. Similarly, among the two written domains, verbs of existence and relationship are more characteristic of academic writing and fulfil largely structural functions. By contrast, mental verbs describe activities and states experienced by humans, including cognitive, perceptual, and affective meanings. This category of verbs occurs more frequently in news discourse than in academic prose, making them particularly suitable for analysis, as they offer a more productive domain for observing stylistic variation in how information is presented and interpreted across different systems. The dominant sub-category in terms of frequency is that of mental verbs describing cognitive states, which are here labelled "stative cognitive verbs".

Following pre-processing of the ST, all verb lemmas contained in the corpus were extracted using a custom Python script to create a frequency-based wordlist. From the corpus used for this

study, concordance lines were extracted for the three most frequent stative cognitive verbs shared across the two news subgenres, namely *know*, *think* and *believe*. These three verbs appear particularly interesting as their semantic generality provides a 'blank canvas' for lexical choice; because a single English verb can be mapped to a wide array of Italian equivalents and reformulations: i.e., *sapere*, *conoscere*, *intendersi di* for *know*; *pensare*, *credere*, *ritenere*, and *trovare* + adjective for *think*, and *credere*, *pensare*, *ritenere*, *secondo* + pronoun for *believe*. Moreover, some Italian equivalents are shared across the three English verbs, resulting in potential interchangeability in translation. Therefore, these verbs might provide interesting insights into the stylistic flexibility and lexical variety of the tools used for translations, assessing their potential for over-generalization or reformulation.

For each occurrence of these verbs, the corresponding translation equivalents were identified from concordance lines extracted from the TT corpora produced by both systems. The translation equivalents were subsequently annotated in a spreadsheet with part-of-speech information, and quantified to enable a systematic comparison of lexical variation across MT systems and news subgenres.

Results

Across the corpus, the three selected verbs occur 1,110 times in total (*know* = 478; *think* = 430; *believe* = 202). Semi-automatic sentence alignment issues excluded some concordances from the analysis; therefore, only 1,073 concordance lines were successfully analyzed (*know* = 454; *think* = 431; *believe* = 199). Translation equivalents are reported only for concordance lines in which the target verbs were successfully identified as such, due to occasional lemmatization and part-of-speech tagging issues in the automatic extraction process. In addition, instances in which the verbs formed part of proper names, such as song titles or movement names, were excluded from the analysis, as they were all retained in English. Based on these considerations, the frequencies of the three verbs are distributed as displayed in Table 2:

Verb	News reports	Opinion articles	Total frequency
know	268	184	452
think	196	227	423
believe	104	94	198
			1,073

Table 2. Distribution of the three selected verbs across the ST corpus.

Table 3, Table 4 and Table 5 present a frequency-based comparison of translation equivalents produced by the two MT tools (Google Translate and ChatGPT-5.2) across opinion articles and news reports, focusing on three high-frequency verbs (*know*, *think*, *believe*). ST frequencies of the verbs, as well as relative frequencies (RF) of the translation equivalents, are reported in the three tables.

The analysis reveals differences in how GT and GPT handle the translation of the three cognitive verbs across genres. A first general observation concerns the distribution of core equivalents. Across all datasets, both systems opt for the same translation equivalents, particularly *sapere* for *know*, *pensare* for *think* and *credere* and *ritenere* for *believe*.

For the verb *know* (Table 3), in opinion articles, both systems rely primarily on *sapere*, *noto*, and *conoscere*, but GT shows a wider variety of low-frequency items. GPT, on the other hand, maintains a tighter distribution while also introducing functionally equivalent paraphrases such as the verbal phrases *avere notizia* (literally 'to have news on', in this context 'to hear from [someone]') and *non avere idea* ('to have no idea'). In news reports, GT again displays greater variety, including more reformulations.

know		News reports (n = 268)						Opinion articles (n = 184)			
GT		**RF**	**GPT**		**RF**	**GT**		**RF**	**GPT**		**RF**
sapere	V	[0.664]	sapere	V	[0.653]	sapere	V	[0.663]	sapere	V	[0.459]
conoscere	V	[0.134]	noto	ADJ	[0.138]	conoscere	V	[0.158]	conoscere	V	[0.168]
noto	ADJ	[0.123]	conoscere	V	[0.134]	noto	ADJ	[0.082]	noto	ADJ	[0.092]
conosciuto	ADJ	[0.049]	conosciuto	ADJ	[0.045]	conosciuto	ADJ	[0.022]	conosciuto	ADJ	[0.022]
capire	V	[0.007]	capire	V	[0.015]	capire	ADJ	[0.016]	capire	V	[0.016]
chissà	ADV	[0.007]	chiamare	V	[0.007]	riconoscere	V	[0.011]	chissà	ADV	[0.011]
scoprire	V	[0.004]	chissà	ADV	[0.004]	chissà	ADV	[0.011]	avere notizia	PHR	[0.005]
ignorare	V	[0.004]	individuare	V	[0.004]	conoscenza	N	[0.005]	definire	V	[0.005]
consapevolezza	N	[0.004]				risaputo	ADJ	[0.005]	non avere idea	PHR	[0.005]
essere a conoscenza	PHR	[0.004]				consapevolezza	ADJ	[0.005]	*omission*		[0.005]
						chiamare	V	[0.005]			
						essere a conoscenza	PHR	[0.005]			
						consapevole	N	[0.663]			
						omission		[0.663]			

Table 3. Translation equivalents for the verb *know* with GT and GPT across news reports and opinion articles.

Overall, *think* (Table 4) displays the greatest dispersion in both genres, and it is therefore the verb that lends itself more to a more varied type of translation. In news reports, *think* is translated predominantly as *pensare* in both systems, with very similar frequencies. Differences emerge in the distribution of secondary options. GT shows a preference for verbal alternatives such as *ritenere*, *credere*, and *considerare*, along with other lower-frequency verbs. On the other hand, in GPT, the main equivalents *credere* and *ritenere* are followed by more phraseological or interpretive renderings, such as *secondo* + subject (literally 'according to…'), *avere dubbi* ('to have doubts'), and *si direbbe* ('one would say'). As for opinion articles, the main equivalent for *think* is *pensare*, with GPT showing a higher frequency (0,806 vs. 0,744 RF in GT). In contrast with news reports, here GT displays a wider range of lexical and phraseological alternatives, including *secondo* + subject, *venire in mente* ('to come to mind'), *trovare* + adjective (literally 'to find [sth] …', meaning 'to think it is …"), and per *come la vedo io* ('the way I see it'), suggesting greater lexical dispersion also at the level of reformulations if compared to GPT.

The verb *believe* (Table 5) is the least variable item with the least pronounced differences between systems. Both GT and GPT rely predominantly on *credere* and *ritenere* in both genres. In the GPT equivalents for news reports, while credere reaches a RF of 0,615 the second most frequent option, *ritenere*, drops to 0,375. On the other hand, in GT, the difference between the frequencies of the same two equivalents is not so marked (*credere* 0,528 vs. *ritenere* 0,423). The data therefore suggest that GPT relies more heavily on a single dominant verb, whereas GT maintains a more balanced use of available alternatives, distributing frequencies across a wider range of alternatives. Independent of the two genres, both systems include reporting-style phrases such as *a suo avviso* ('in his/her view'), and *secondo* + subject in GT, and *che piaccia o no* ('like it or not') in GPT for opinion articles.

think		News reports (n = 196)				Opinion articles (n = 227)				
GT		**RF**	**GPT**		**RF**	**GT**		**RF**	**GPT**		**RF**
pensare	V	[0.775]	pensare	V	[0.796]	pensare	V	[0.744]	pensare	V	[0.806]
credere	V	[0.071]	credere	V	[0.086]	credere	V	[0.114]	credere	V	[0.114]
ritenere	V	[0.082]	ritenere	V	[0.056]	consid-erare	V	[0.04]	ritenere	V	[0.441]
consid-erare	V	[0.03]	secondo + subject	PHR	[0.015]	ritenere	V	[0.035]	trovare + adjective	V	[0.013]
riflettere	V	[0.01]	considerare	V	[0.01]	riflettere	V	[0.013]	consid-erare	V	[0.009]
trovare + adjective	V	[0.01]	trovare + ad-jective	V	[0.01]	secondo + subject	PHR	[0.009]	riflettere	V	[0.004]
stimare	V	[0.005]	sbagliarsi	V	[0.005]	venire in mente	PHR	[0.009]	venire in mente	PHR	[0.004]
sbagliarsi	V	[0.005]	illudersi	V	[0.005]	trovare + adjective	V	[0.009]	chiedersi	V	[0.004]
illudersi	V	[0.005]	avere dubbi	PHR	[0.005]	affrontare	V	[0.004]	con-vinzione	N	[0.004]
omission		[0.005]	si direbbe	PHR	[0.005]	immagi-nare	V	[0.004]			
						concepire	V	[0.004]			
						chiedersi	V	[0.004]			
						per come la vedo io	PHR	[0.004]			
						chiedersi	V	[0.004]			
						omission		[0.004]			

Table 4. Translation equivalents for the verb *think* with GT and GPT across news reports and opinion articles.

believe		News reports (n = 104)				Opinion articles (n = 94)					
GT		**RF**	**GPT**		**RF**	**GT**		**RF**	**GPT**		**RF**
credere	V	[0.528]	credere	V	[0.615]	credere	V	[0.680]	credere	V	[0.723]
ritenere	V	[0.423]	ritenere	V	[0.375]	ritenere	V	[0.255]	ritenere	V	[0.244]
a suo av-viso	PHR	[0.019]	secondo + sub-ject	PHR	[0.009]	secondo + subject	PHR	[0.021]	secondo + subject	PHR	[0.01]
pensare	V	[0.009]				a mio/suo avviso	PHR	[0.021]	pensare	V	[0.01]
secondo + subject	PHR	[0.009]				convinto	ADJ	[0.021]	che piaccia o no	PHR	[0.01]
convinto	ADJ	[0.009]									

Table *5*. Translation equivalents for the verb believe with GT and GPT across news reports and opinion articles.

As regards structural variation in the translation strategies adopted by the two systems (Table 6), the data show that GT more frequently preserves a direct verb-based (V-V) correspondence (GT 16 vs. GPT 14) across both genres. At the same time, GT demonstrates a higher tolerance for morphological variation, including nominalization (verb to noun; V-N), adjectival (verb to adjective, V-ADJ) and adverbial forms (verb to adverb; V-ADV). While these forms are not absent in GPT, they tend to occur less frequently and with lower diversification. For instance, it is interesting to point out that GPT uses nominalization only once across both genres. By contrast, it more readily introduces phraseological reformulations. This is visible in the use of prepositional phrases (PP) (*secondo + subject*, *a suo avviso*) and clausal constructions (CLAUSE) (*si direbbe*, *che piaccia o no*), particularly in the translation of *think*; meanwhile, GT produces a less varied output in terms of phraseological reformulations, with the most frequent of them being the construction *secondo + subject*.

	News reports		Opinion articles	
	GT	GPT	GT	GPT
V-V	16	14	16	14
V-ADJ	4	3	5	2
V-N	1	0	2	1
V-ADV	1	1	1	1
V-PP/CLAUSE	3	4	5	5

Table 6. Structural variation of translation strategies, measured as the number of distinct translation equivalents provided by each MT system across the two news genres.

Moreover, some differences with respect to genre conventions might also be observed. In news reports, both systems show a strong preference for reporting verbs such as *sapere*, *credere*, *ritenere*, and *pensare*, with limited variation and a high degree of overlap between GT and GPT outputs. Therefore, the data might be interpreted as reflecting the informational and objective style of the news report genre. In opinion articles, by contrast, the output is more varied with regard to reporting verbs, also including more word-class transformation and the presence of phrasal or clausal constructions. The use of a wider range of lexical items and more frequent fixed phrases and expressions can be observed in the output of both systems, such as for *per come la vedo io*, *a suo avviso*, or *venire in mente*. Variation increases particularly in the GT output for *think*, in which many more equivalents are provided compared to GPT.

Overall, these results point to minor differences in MT output. GT shows lexically diverse equivalents, often relying slightly more on word-class transformation. GPT, by contrast, exhibits a more interpretive style, favouring reformulations while reducing lexical dispersion. Genre conventions seem to be reflected in the output of both systems, with opinion articles having a wider range of equivalents and especially phrasal and clausal reformulations, which refer to the more argumentative structure of the ST.

4 Conclusion

This study has examined lexical variation in English–Italian MT by analyzing the distribution of translation equivalents for three high-frequency stative cognitive mental verbs in outputs produced by GT and GPT by answering the following research questions:

(1) How do GT and GPT differ in lexical variation when translating English mental cognitive verbs - *know*, *believe*, *think* –into Italian?;

(2) Does genre play a role in the extent to which these stylistic tendencies differ?

The results show that the two systems differ slightly in their stylistic realization, with GT exhibiting greater sensitivity to morphological variation and producing a wider range of translation equivalents. GPT, by contrast, tends to rely on a more restricted set of equivalents, resulting in a more lexically concentrated output and a wider range of phrasal and clausal reformulations. While GT tends to distribute meaning across a broader set of lexical choices, GPT appears to achieve variation more frequently through restructuring at the phrase and clause level rather than through expansion of lexical alternatives. Furthermore, the results might suggest some sensitivity to news genre: opinion articles display greater lexical and structural variation than news reports in both systems, indicating that textual conventions present in the ST might be relevant enough to shape the range of translation choices adopted by the systems. This is particularly evident with the equivalents of the verb *think*, which show the highest degree of variation, while *believe* remains comparatively stable across both systems.

This study has several limitations. First, the analysis is restricted to a small set of verbs, which, albeit frequent and representative of stative cognitive meaning, cannot capture the full range of lexical behavior in MT. Second, technical constraints might have influenced the results, such as the parameters set in the Python script when performing the API and in the prompt. Third, the focus on quantitative measures of lexical variety does not account for contextual or discourse-level factors that may influence translation choices.

Future work within the broader research project will focus on expanding the TT corpora with the translations performed by other NMT systems and LLMs to assess whether their outputs provide similar results and whether the results obtained can be generalized. Subsequently, the TT corpora will be compared with a comparable Italian news corpus. Moreover, the analysis could be integrated by examining a broader range of lexical items and performing a qualitative analysis of concordance lines, which would allow for a more detailed investigation of how lexical choices interact with context and discourse. Furthermore, including other indicators of variation such as syntactic complexity, could provide a more comprehensive account of stylistic variation in MT.

The findings of this study contribute to existing research on machine translation for this specific language combination, which has not been extensively researched until now. They provide a quantitative account of stylistic variation in English-Italian translations and offer a comparative perspective on NMT and LLM outputs, showing how different systems and genres interact to produce lexical variation.

References

Biber, D., Johansson, S., Leech, G., Conrad, S., and Finegan, E. Longman Grammar of Spoken and Written English.

Bowker, L. (2024). Risks for lay users in machine translation and machine translation literacy. In The Social Impact of Automating Translation (pp. 60-76). Routledge.

Bullard, G. (2023). Smart ways journalists can exploit artificial intelligence. Nieman Reports. https://niemanreports.org/artificial-intelligence-newsrooms/ (niemanreports.org). Last accessed:2026-05-26.

Cagnan, P. (2025). Uso dell'intelligenza artificiale nelle redazioni italiane. Ordine dei Giornalisti. https://www.odg.it/uso-dellintelligenza-artificiale-nelle-redazioni-italiane/59863. Last accessed:2026-05-26.

Caswell, David (2019). Structured Journalism and the Semantic Units of News. Digital Journalism, 7(8), 1134-1156.

Cools, H., and Diakopoulos, N. (2024). Uses of Generative AI in the Newsroom: Mapping Journalists' Perceptions of Perils and Possibilities. Journalism Practice, 1–19. https://doi.org/10.1080/17512786.2024.2394558

Daems, J. (2022). Dutch literary translators' use and perceived usefulness of technology: The role of awareness and attitude. In Using technologies for creative-text translation (pp. 40-65). Routledge. doi: 10.4324/9781003094159-3

Daems, J., De Clercq, O., and Macken, L. (2017). Translationese and post-editese: How comparable is comparable quality? Linguistica Antverpiensia, New Series: Themes in Translation Studies, 16, 89–103. doi: 10.52034/lanstts.v16i0.434.

De Clercq, O., De Sutter, G., Loock, R., Cappelle, B., and Plevoets, K. (2021). Uncovering machine translationese using corpus analysis techniques to distinguish between original and machine-translated French. Translation Quarterly, 101, 21–45. ⟨hal-03406287⟩.

Espinosa-Giménez, M., and Forcada, M. L. (2024). Lexical impoverishment in neural machine translation: automatic and human study of Spanish-English noun translation. Tradumàtica tecnologies de la traducció, (22), 162-180. doi: 10.5565/rev/tradumatica.428.

Hermand, M. H. (2014). The Euroregional discourse. Increasing evidence towards legitimizing an institutional space. Mots. Les langages du politique, (106). doi: 10.4000/mots.21839.

Honnibal, M., Montani, I., Van Landeghem, S., and Boyd, A. (2020). spaCy: Industrial-strength Natural Language Processing in Python. https://doi.org/10.5281/zenodo.1212303

Jiang, Y., and Niu, J. (2022). A corpus-based search for machine translationese in terms of discourse coherence. Across Languages and Cultures, 23(2), 148-166. doi: 10.1556/084.2022.00182.

Kocmi, T., Artemova, E., Avramidis, E., Bawden, R., Bojar, O., Dranch, K., Dvorkovich, A., Dukanov, S., Fishel, M., Freitag, M., Gowda, T., Grundkiewicz, R., Haddow, B., Karpinska, M., Koehn, P., Lakougna, H., Lundin, J., Monz, C., Murray, K., Nagata, M., Perrella, S., Proietti, L., Popel, M., Popović, M., Riley, P., Shmatova, M., Steingrímsson, S., Yankovskaya, L., Zouhar, V. (2025). Findings of the WMT25 General Machine Translation Shared Task: Time to Stop Evaluating on Easy Test Sets. In B. Haddow, T. Kocmi, P. Koehn, and C. Monz (Eds), Proceedings of the Tenth Conference on Machine Translation (pp. 355–413). Association for Computational Linguistics. doi: 10.18653/v1/2025.wmt-1.22.

Kong, D., and Macken, L. (2025a). Can Peter Pan Survive MT? A Stylometric Study of LLMs, NMTs, and HTs in Children's Literature Translation.

Kong, D., and Macken, L. (2025b). Decoding Machine Translationese in English-Chinese News: LLMs vs. NMTs. In 20th Machine Translation Summit.

Loock, R. (2020). No more rage against the machine: How the corpus-based identification of machine-translationese can lead to student empowerment. Journal of Specialised Translation, (34), 150–170. ⟨halshs-02913980⟩.

Macken, L., De Clercq, O., Paulussen, H. (2011). "Dutch Parallel Corpus: A Balanced Copyright-Cleared Parallel Corpus." META 56 (2): 374–90.

Molino, A. (2011). A Corpus-Based Study of Stance Adverbials in Learner Writing. Roma: Aracne Editrice.A

Murru, M. F., and Carlo, S. (2024). AI e newsmaking: Un'indagine esplorativa nelle redazioni nazionali e locali italiane. Mediascapes Journal, 23(1), 184–198. https://rosa.uniroma1.it/rosa03/mediascapes/article/view/18714

Murtaza, G., and Qasmi, N. Q. (2013). Style and stylistics: An overview of traditional and linguistic approaches. Galaxy: international multidisciplinary research journal, 2(3), 2278-9529.)

Ngai, J. (2022). Evaluation across newspaper genres: Hard news stories, editorials and feature articles. Routledge. doi: 10.4324/9781003150640.

Niu, J., and Jiang, Y. (2024). Does simplification hold true for machine translations? A corpus-based anal-

ysis of lexical diversity in text varieties across genres. Humanities and Social Sciences Communications, 11(1), 1-10. doi: 10.1057/s41599-024-02986-7.

Roush, Chris. (2023). What Reuters is telling its journalists about using artificial intelligence. Talkinbiz. https://talkingbiznews.com/?s=What+Reuters+is+telling+its+journalists+about+using+artificial+intelligence. Last accessed:2026-05-26.

The New York Times Trust Team. (2024). How The New York Times uses AI in journalism. The New York Times. https://www.nytimes.com/2024/10/07/reader-center/how-new-york-times-uses-ai-journalism.html. Last accessed:2026-05-26.

Vanmassenhove, E., Shterionov, D., and Gwilliam, M. (2021). Machine translationese: Effects of algorithmic bias on linguistic complexity in machine translation. In Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume (pp. 2203–2213). Association for Computational Linguistics. doi: 10.18653/v1/2021.eacl-main.188.

Vanmassenhove, E., Shterionov, D., and Way, A. (2019). Lost in Translation: Loss and Decay of Linguistic Richness in Machine Translation. In Proceedings of Machine Translation Summit XVII Volume 1: Research Track (pp. 222-232). European Association for Machine Translation.

Webster, R., Fonteyne, M., Tezcan, A., Macken, L., and Daems, J. (2020). Gutenberg Goes Neural: Comparing Features of Dutch Human Translations with Raw Neural Machine Translation Outputs in a Corpus of English Literary Classics. Informatics, 7(3), 32. doi:10.3390/informatics7030032.