European Journal of Multilingualism and Translation Studies
ISSN: 2734 - 5475
ISSN-L: 2734 - 5475
Available on-line at: www.oapub.org/lit

DOI: 10.46827/ejmts.v6i1.715 Volume 6 │ Issue 1 │ 2026

AI-BASED TRANSLATION QUALITY FOR LOW-RESOURCE LANGUAGES: THE CASE OF MALTESE

Sergio Portelli[i]
Professor,
Department of Translation, Terminology and Interpreting Studies
Faculty of Arts,
University of Malta,
Malta

Abstract:
Maltese is the national language of the Republic of Malta and shares its official language status with English. Despite being used by a small community mostly in oral communication and in a few formal domains, it became an official language of the European Union in 2004. In the last two decades, because of EU membership, translation requirements in Maltese have greatly increased. However, the relative scarcity of Maltese texts in many domains, especially in technical fields, has created a critical data deficiency for training large language models (LLMs). Hence, the quality of AI-driven translation for Maltese is generally perceived to be inadequate. However, to date, no studies have been made to assess translation quality related to the use of AI-based technologies for the Maltese language. To address this research gap, the present small-scale study evaluates the performance of two prominent AI-based translation tools, Google Translate and ChatGPT, on a 6000-word corpus of 20 texts translated from Italian into both Maltese and English. The raw output was systematically evaluated using an adapted DQF-MQM error typology template. The results show that in the case of Maltese, Google Translate made almost three times more errors with respect to English, while ChatGPT generated over seven times the errors for Maltese. The analysis concludes that despite the high status of Maltese in the EU's multilingual setting, the limitations of Maltese as a low-resource language still persist, and a highly cautious approach must be taken by Maltese translators and post-editors when using AI-based tools for translation.

Keywords: Maltese, artificial intelligence, translation, error analysis

1. Introduction

Maltese is the national language of the Republic of Malta, an archipelago of approximately 300 km^2 in the central Mediterranean, around 100 km south of Italy and 350 km north of Libya. The island state, which has been a member of the European Union since 2004, has two official languages, Maltese and English, and prior to accession successfully requested that the former

[i] Correspondence: email sergio.portelli@um.edu.mt

be recognised as an official language of the EU. Consequently, within the multilingual framework of the European institutions, Maltese is at par with other 23 official languages spoken by many millions of people. It has an estimated total of 530,000 speakers in Malta and among the diaspora (Ethnologue 2025), and has long co-existed with English, which is the language of education and mostly used in written communication (National Council for the Maltese Language 2021: 4). For this reason, the relatively limited number of updated dictionaries and digital texts in Maltese for many technical and scientific domains puts the language in the category of low-resource languages, defined by Cieri et al. as languages 'that have fewer technologies and especially data sets relative to some measure of their international importance' (2016: 4545). Despite its importance as an official EU language, Maltese faces a disadvantage because of the limited datasets that exist on which to train large language models (LLMs), thus hampering automatic translation quality. The present small-scale study is the first to carry out a quality-based assessment of AI-based translation for the Maltese language, and seeks to provide some data on current limitations that have implications on user experience, post-editing effort and translator training.

2. AI and Translation

In the mid-2010s, AI-powered Neural Machine Translation (NMT), characterized by deep learning algorithms trained on massive datasets significantly improved translation accuracy and efficiency, especially for common language pairs and straightforward texts, by dealing better with the nuances of language (Mohamed et al. 2024: 25564). Machine translation (MT) tools like Google Translate, Bing, Microsoft Translator and DeepL are widely used by general internet users (Al Romany & Kadhim 2024, Moneus & Sahari 2024), and Google Translate alone handles around 143 billion words per day (Eszenyi et al. 2023). In November 2022, chatbot-based GPT AI tools were made available to internet users, and became fast and cost-effective translation tools accessible to individuals and small businesses globally (Moneus & Sahari 2024; Shahmerdanova 2025).

Despite these rapid advancements, AI translation still faces significant challenges, particularly when dealing with low-resource languages (LRLs). AI still struggles to capture cultural nuances, idiomatic expressions, and the translation of complex or ambiguous contexts, which hampers overall translation quality (Shahmerdanova 2025). This is especially the case when creative texts are involved. Karabayeva & Kalizhanova (2024) noted that elaborate rhetorical features, allusions, covert culture-specific references and individual style pose significant difficulties for AI, and that while ChatGPT and DeepL can convey literal meanings, they struggle with nuance, fluidity, metaphor, alliteration, rhyme, imagery, and tone. On their part, AI translations of legal texts can overlook the legal implications of specific phrases and the legal effect of specific words (Al Romany & Kadhim 2024; Moneus & Sahari 2024). Moneus & Sahari (2024) found that human translators consistently score higher in accuracy, competency, and the appropriate use of legal language and tone in legal translations between English and Arabic.

The emergence of LLMs has brought about a significant improvement in MT and natural language processing, resulting in a remarkable translation quality for high resource languages (HRLs). However, these models face serious challenges when applied to LRLs, especially when the language pair consists of two LRLs (Hendy et al. 2023). Whereas wider-used languages such as English, French, Spanish and Chinese provide datasets for the training of LLMs that are large enough to result in high-quality translation output, LRLs lack sufficient online presence and parallel texts to reach the same quality. In order to mitigate this scarcity of resources, various efforts are being made; to address the issue of data scarcity for European LRLs, in July 2025 Microsoft announced the launch of a collaborative project with the University of Strasbourg to increase the availability of multilingual data in languages with 'relatively low representation in online content', such as Estonian, Alsatian, Slovak, Greek, and Maltese (Smith 2025).

From a translation perspective, strategies are being devised to increase data supplies, especially for Neural Machine Translation (NMT) engines. Back translation has yielded positive results for LRL-HRL pairs, including Nepali-English and Sinhala-English by using target monolingual data to create synthetic parallel data, often increasing the corpus size fivefold, while the use of Transfer Learning demonstrated an improved performance over baseline systems for languages such as Khmer-English and Pashto-English (Goyle et al. 2023). Limited resource availability also affects the assessment of LLM performance on LRLs, due to the difficulty of widely-used metrics like BLEU to capture the actual quality of LLM outputs. Although other techniques are being explored, a cautious approach to the results must be adopted, taking into consideration the inferior quality of the data available for LRLs (Hendy et al. 2023).

3. The scope of the study

This study aims to investigate the performance of an NMT tool (Google Translate) and a GPT AI chatbot (ChatGPT 5) in translating a corpus of five texts each for four text types from Italian, a HRL, to Maltese, a LRL, for a total of approximately 6000 words. Performance is assessed from a user's perspective, i.e. on the number of interventions a post-editor would have to make to bring the raw output to publishable quality. The frequency of interventions has an impact on time, cost and sustainability, which are all factors that must be considered when choosing between a human-aided machine translation and a human translation for high-quality output. The quantitative aspect of the study is complemented by a qualitative component in that the translation errors are classified according to a customised version of the Dynamic Quality Framework (DQF) and the Multidimensional Quality Metrics (MQM) error typology. The metric, which is a model used by major stakeholders such as Google Translate and Microsoft Translator, is language-independent and adaptable to different domains, making it suitable for multilingual programs (Dranch n.d.). The classification of errors is carried out to provide a picture of the most frequent errors made by the tools in the different text types. However, given the scope of the study, the weighting of error types was not taken into consideration. The same process is carried out with English as a target language for comparative purposes, in order to assess the difference in raw output quality between English and Maltese, a HRL and a LRL respectively. The findings are presented according to text types and are discussed according to Maltese user requirements and their implications on translator training.

4. Methodology

The study is based on a corpus of 20 texts in Italian of approximately 300 words each pertaining to four domains: literature, news, finance and medicine. The domains were chosen according to their different registers. Literature and news texts in Italian are characterised by a frequent use of metaphors and idiomatic expressions, with the former more strongly dependent on the author's personal style. On the other hand, finance and medicine are technical domains that focus on information, follow conventional styles and contain standard terminology. Five texts that had never been translated into English and Maltese were chosen for each domain. This way, a comparison between the performance of Google Translate and ChatGPT for each domain and for both target languages could be carried out. The 300-word limit was established to strike a balance between the stylistic and terminological representativeness of the passage with respect to the specific text and the manageability of the small-scale study. Although debatable, given its limited dataset, this approach was deemed appropriate given the comparative nature of the study.

Google Translate was the NMT chosen for the study because it supports Maltese and is one of the most widely used online translation tools in the world with over 500 million users (Bhattacharya 2026). ChatGPT was chosen because it is the most popular general purpose chatbot AI, given that its market share stands at above 70% (alltopeverything.com 2025). The translations were initially done using version 4o ('omni'), but after OpenAI launched ChatGPT 5 in May 2025, they were redone to ensure that the findings remained relevant according to the latest version on the market.

As mentioned above, a customised version of the DQF-MQM typology was used to classify the errors that would have required the intervention of a post-editor. The taxonomy adopted is shown in Table 1 below:

Table 1: Taxonomy of errors adapted from the DQF-MQM error typology (TAUS Dynamic Quality Evaluation)

High-level error type	Granular error type	Definition
Accuracy	Mistranslation	The target content does not accurately represent the source content.
	Untranslated text	Content that should have been translated has been left untranslated.
Fluency	Spelling	Issues related to spelling of words.
	Grammar	Issues related to the grammar or syntax of the text, other than spelling and orthography.
Terminology		A term (domain-specific word) is translated with a term other than the one expected for the domain or otherwise specified.
Style	Unidiomatic	The content is grammatical, but not idiomatic.
Design		There is a problem relating to design aspects (vs. linguistic aspects) of the content.

Verity	Culture-specific reference	Content inappropriately uses a culture-specific reference that will not be understandable to the intended audience.

The texts were selected on the grounds that no English translation had been published in any format before the research. Each text was given a code that consisted of the first three letters of the domain and a number. The literary texts were taken from four novels and a collection of short stories, namely, *La distrazione*, by Luciano De Crescenzo (Mondadori 2000, LIT 1), *Mandami a dire*, by Pino Roveredo (Bompiani 2005, LIT 2), *La casa del padre* by Giorgio Montefoschi (Bompiani 1994, LIT 3), *Carossa*, by Claudio Marabini (Rizzoli 1990, LIT 4), and *Il coraggio del pettirosso*, by Maurizio Maggiani (Feltrinelli 1995, LIT 5). All the news texts were taken from online portals: rainews.it on Italian economic statistics (NEW 1), quifinanza.it on the Net Zero Banking Alliance (NEW 2), corriere.it on multinationals selling their food brands (NEW 3), ilgiornaledellarte.it on a new archaeological site opening to the public in Rome (NEW 4), and avvenire.it on the rise of food prices due to climate change (NEW 5). The financial texts were all found online: a blog on Quarterly Business Review models on clickup.com (FIN 1), a text on the governance of Italian companies on consob.it (FIN 2), the 2025 financial report of TIM Group on gruppotim.it (FIN 3), a text on business accounting on affarifinanza.it (FIN 4), and another on the relationship between banks and companies on fondazionenazionalecommercialist.it (FIN 5). Lastly, the texts form the medical domain were taken from the following websites: topdoctors.it on kidney failure (MED 1), policlinico.unina.it on surgical processing (MED 2), epc.it on occupational medicine (MED 3), chirurgiadeilinfatici.com on types of surgical procedures (MED 4), and msdmanuals.com on endoscopy (MED 5). The variety of texts was intended to give insights into the translation quality provided by the two tools in Maltese and to allow a reliable comparison with the quality of the raw output in English.

Since Maltese is a low-resource language, the raw output was annotated manually by the researcher to guarantee reliability and consistency. A table listing the number of errors for each error type in Table 1 was produced for each text, for Google Translate and ChatGPT and for English and Maltese. Additionally, a table was produced for each text type containing the total number of errors in the respective domain. The format of the tables can be seen in Table 2:

Table 2: Template used for error classification

High-Level Error Type	Granular Error Type	Google EN	ChatGPT EN	Google MT	ChatGPT MT
Accuracy	Mistranslation				
	Untranslated text				
Fluency	Spelling				
	Grammar				
Terminology	Terminology				
Style	Unidiomatic				
Formatting	Formatting				
Verity	Culture-specific reference				
Total	**Total**				

Each text was translated four times, twice in English and twice in Maltese. For ChatGPT, a zero-shot prompt was used since multiple prompts would have possibly improved the output but the translation process would have taken more time and effort on the part of the post-editor. The prompt used was: 'Translate the following text in the domain of [literature/news/finance/medicine] from Italian into [English/Maltese] using the appropriate register in the target language'. The domain was specified in the prompt to provide context and allow the tool to identify the parameters for choosing the appropriate register and terminology. Each translation was only carried out once for the same reasons of time and effort mentioned above.

For each text, the four raw outputs were analysed and the translation errors were highlighted and classified accordingly. Since the scope of the study was based on post-editor intervention frequency, repeated errors were counted as separate instances of required intervention. The numbers were then inserted in the corresponding table. At the end of the process, four tables with totals were produced, one for each domain.

5. Findings and Analysis

5.1 Literary texts

The results for the literary texts are shown in Table 3 below:

Table 3: Error frequency in raw output of literary texts

High-Level Error Type	Granular Error Type	Google EN	ChatGPT EN	Google MT	ChatGPT MT
Accuracy	Mistranslation	12	4	29	21
	Untranslated text	1	0	3	0
Fluency	Spelling	0	0	5	2
	Grammar	13	2	36	31
Terminology	Terminology	1	0	10	16
Style	Unidiomatic	6	1	13	25
Formatting	Formatting	1	0	1	0
Verity	Culture-specific reference	2	0	2	0
Total	**Total**	**36**	**7**	**99**	**95**

The numbers clearly show that both Google Translate and ChatGPT perform much better in English than in Maltese. While this is not surprising given the difference in data availability for LLM training between the two languages, the extent of the difference is noteworthy. Post-editing effort required for the Google Translate output is three times higher for Maltese, and thirteen times higher for the ChatGPT versions. It is also interesting to note that whereas ChatGPT performs much better than Google Translate in English, the numbers for both in Maltese are very similar. Consequently, for the texts concerned, using chatbot AI did not bring any significant advantages over machine translation as regards Maltese. Google Translate performed best in LIT 5 (15 errors) and worst in LIT 2 (26), whereas ChatGPT was more consistent with an average of 18 errors per text.

The main issues for Maltese were mistranslations, grammar, style and terminology. Given that the creative nature of literary works may have an impact on meaning due to allusions, metaphors and other rhetorical devices that require deeper pragmatic interpretation, mistranslations may be expected. This is confirmed, albeit at a lower frequency, by the occurrences in English. Grammar issues were mostly of a syntactical nature, where the raw output occasionally tended to follow the source text too closely. However, the use of wrong pronouns was also encountered, especially when the source text did not specify the subject explicitly. This is not uncommon in Italian and other neo-Latin languages. Gender mismatch or inconsistency was also noted where the Italian word and its Maltese counterpart are identical but have different genders, e.g. 'sistema' (system), which is grammatically masculine in Italian and feminine in Maltese. Style issues were mostly related to a wrong choice of register, since in certain cases the Maltese output oscillated between the appropriate register (e.g. conversational or informal) and a more formal one. This, too, is unsurprising considering that formal texts are pre-eminent in the Maltese datasets found in corpora such as the Korpus Malti (MLRS 2024).

Terminology errors were also to be expected, since Maltese lacks terms in many domains (Portelli 2025). However, errors were recorded even where Maltese terms were available, such as in the case of 'vasca da bagno', which was rendered by Google Translate in the Maltese version as 'bathtub', thus using the English term instead of the Maltese equivalent 'banju'. In other cases, terms were kept in Italian both by Google Translate and ChatGPT, such as 'casamento' ('residential block'). The way both tools handled terminology was particularly interesting, since they resorted to using Italian or English terms when they could not find a Maltese equivalent. Since the study used Italian source texts, the tools occasionally kept the word in the source when faced with a lack of equivalents; however, the option to resort to English is an indication that the presence of English words in the Maltese dataset possibly directed the tools towards that option.

5.2 News texts

The results for the news texts are shown in Table 4 below:

Table 4: Error frequency in raw output of news texts

High-Level Error Type	Granular Error Type	Google EN	ChatGPT EN	Google MT	ChatGPT MT
Accuracy	Mistranslation	4	1	10	7
	Untranslated text	0	0	0	0
Fluency	Spelling	0	0	0	4
	Grammar	9	3	15	15
Terminology	Terminology	0	1	7	7
Style	Unidiomatic	5	2	16	14
Formatting	Formatting	0	2	0	2
Verity	Culture-specific reference	1	0	1	0
Total	**Total**	**19**	**9**	**49**	**49**

In the news text, Maltese translations, post-editing requirement was significantly less than for the literary ones. This may be due to the stronger stylistic uniformity in the domain of journalism and the large proportion of news texts in Maltese present on the internet with respect to texts in other domains, thus affecting the composition of the datasets and producing better translation results. The tools registered 50% less errors, but the numbers were still considerably higher than for the English translations. The most common errors were related to mistranslations, grammar, terminology and style, which confirms the areas where the tools find tougher challenges to overcome. Google Translate performed best in NEW 5 (4 errors) and worst in NEW 1 (14). On the other hand, ChatGPT made the fewest errors in NEW 1 (6) and the most in NEW 4 (17). The similar overall quality of the performance of Google Translate and ChatGPT in Maltese confirmed that both tools struggled to render narrative texts that use metaphors, figurative language, rhetorical devices and uncommon terms.

The most interesting errors were of a terminological nature. Google Translate occasionally resorted to English instead of using the Maltese equivalent (e.g. 'biscotti' translated as 'cookies' instead of 'gallettini'). In one case, it interestingly resorted to Spanish when it translated the word 'magnifico' ('magnificent') as 'magnífico', while ChatGPT used 'magnificu', spelled according to nineteenth-century Maltese orthography. Both tools also failed to translate correctly the fraction 'due quinti' ('two fifths'): Google Translate opted for 'żewġ ħamsa', while ChatGPT translated it as 'żewġ minn ħamsa' instead of the established form 'żewġ kwinti'.

Another remarkable terminological error that could potentially reflect a wider phenomenon was the translation of 'carrello della spesa' ('shopping trolley' or 'shopping cart'). It was translated by both tools as 'karrettun tax-xiri', which is a literal translation of 'shopping cart'. The term, which is not used in Maltese and is not found in the Korpus Malti, was found only in non-Maltese websites that automatically use machine translation to change language according to the IP address of the visitor to the site. The term has probably been propagated on the internet by these translations and found itself in the dataset used to train the tools. This unintended process exponentially increases the presence of such terms, which may be found and used by inexperienced translators, thus creating a vicious circle where terminology for low-resource languages such as Maltese, in which term formation may still not be methodologically established, is influenced by the raw output of translation tools or AI chatbots that is published without human quality assurance.

5.3 Financial texts

The results for the financial texts are shown in Table 5 below:

Table 5: Error frequency in raw output of financial texts

High-Level Error Type	Granular Error Type	Google EN	ChatGPT EN	Google MT	ChatGPT MT
Accuracy	Mistranslation	0	0	1	2
	Untranslated text	0	0	0	0
Fluency	Spelling	0	0	3	7
	Grammar	8	2	6	6
Terminology	Terminology	1	0	7	0
Style	Unidiomatic	5	0	15	9
Formatting	Formatting	0	4	0	4
Verity	Culture-specific reference	0	0	0	0
Total	**Total**	**14**	**6**	**32**	**28**

The performance of the tools for the translation into Maltese of the financial texts was better than in the two domains discussed above. In fact, Google Translate made no errors in the translation of FIN 3, although it performed badly in FIN 1 with 21 errors. On the other hand, ChatGPT performed best in FIN 3 (3 errors) and worst in FIN 1 (7). The lesser number of errors in this domain is unsurprising given that technical texts tend to follow a conventional register specific to their respective domain. However, the highest frequency of errors was still registered in style, since the tools occasionally tended to translate the source text too closely at the expense of fluency and naturalness. This was more common in the raw output created by Google Translate, which also performed worse in terminology. ChatGPT did not make any terminological errors in Maltese, suggesting that its much larger dataset produces better results in this domain than the parallel corpora used by Google Translate.

On the other hand, this difference in datasets could be the reason why ChatGPT registered a larger amount of spelling errors than Google Translate. Maltese orthographical rules have changed as recently as 2008, so texts on the internet published prior to that date still follow the old rules. Moreover, texts uploaded after 2008 may still contain obsolete spelling rules depending on the author's awareness or acceptance of the changes. Another contributing factor may be that the percentage of Maltese speakers with advanced writing skills is only 59% (National Statistics Office 2024). Since Malta's internet penetration and social media use amount to 93.5% and 68.7% of the whole population respectively (Kemp 2025), many users who write in Maltese on the internet, especially in their social media posts, are unwittingly feeding Maltese incorrect spelling to the dataset used by generative AI.

As regards the formatting errors that may be noted in both the English and Maltese versions of ChatGPT, these were due to the arbitrary subdivision of the source text in shorter paragraphs and the inclusion of bold type for key concepts. In general, the smaller number of errors in the translation of the financial text type into Maltese indicates that the tools may be more convenient to use for texts in this domain than in the previous two discussed above.

5.4 Medical texts

The results for the medical texts are shown in Table 6 below:

Table 6: Error frequency in raw output of medical texts

High-Level Error Type	Granular Error Type	Google EN	ChatGPT EN	Google MT	ChatGPT MT
Accuracy	Mistranslation	1	0	1	8
	Untranslated text	0	0	0	0
Fluency	Spelling	0	0	2	7
	Grammar	6	0	11	5
Terminology	Terminology	0	0	8	10
Style	Unidiomatic	3	0	9	3
Formatting	Formatting	1	2	1	2
Verity	Culture-specific reference	0	0	0	0
Total	**Total**	**11**	**2**	**32**	**35**

The results for the Maltese translations of medical texts are broadly similar to those of the financial texts. Google Translate was consistent in averaging 5 errors per text, whereas ChatGPT marked a significant difference between its best performance, MED 4 with 2 errors, and MED 2 and MED 5, both with 11 errors each. Most conspicuously, ChatGPT made more translation errors than Google Translate. This was due to ChatGPT's poor performance in the translation of MED 1, where four mistranslations were recorded. As regards terminology, ChatGPT struggled mostly in MED 2 and MED 5, due to a lack of Maltese terms for specific concepts and devices in the specific areas of surgery protocols and endoscopy. In these cases, the tools adopted the same strategies as mentioned above in 5.2. For example, ChatGPT left the term 'prognosi' (prognosis, MED 1) in Italian, instead of using the Maltese equivalent 'pronjosi'. It also formed a calque from the source language for 'lista di presa in carico' (follow-up list, MED 2), creating 'lista ta' presa in karigu' instead of the established equivalent 'lista ta' segwitu'. Interestingly, in both cases, Google Translate used the correct terms. The latter did, however, occasionally resort to Italian, such as when it kept 'visita' (visit, MED 2) instead of using the Maltese 'viżta'.

Google Translate made most errors in grammar and style, mostly due to its tendency to follow more closely the source language's syntactic structures. A noteworthy issue in the tool is that on occasion it misplaced the apostrophe at the end of words by adding them to the beginning of the following word, e.g. 'tista 'tiġi restawrata' instead of 'tista' tiġi restawrata' (it may be restored, MED 1). This error is not done consistently even within the same text, thus creating a post-editing issue that may require more effort and time to identify and correct.

5.5 General performance in percentage terms

The comparison of the data per domain and on the whole corpus between the Maltese and English translations sheds light on the difference in the performance of Google Translate and ChatGPT when translating into the two languages. Table 7 shows the percentage of errors per domain:

Table 7: Total percentage of errors per domain (1500 words each)

	Literary	News	Financial	Medical
Google T. EN	2.4	1.3	1	0.7
ChatGPT EN	0.5	0.7	0.4	0.1
Google T. MT	6.5	3.3	2	2
ChatGPT MT	6.3	3.3	1.9	2.3

Google Translate made three times more errors in Maltese as in English in the literary and news domains, and twice more in the technical ones. On the other hand, the difference in the performance of ChatGPT between the two languages was significantly higher. The AI chatbot consistently produced high quality output in English, remaining well under 1% of errors in all domains. In the medical texts, it performed remarkably well, and the raw output required very little post-editing effort. For Maltese, the accuracy level of the translations was much lower and varied considerably across domains. For literary texts, 6.3% represents more than twelve and a half times more errors than in English, thus requiring a much higher post-editing effort to produce publishable quality target texts. Contrary to English, where ChatGPT performed slightly worse in the news texts than in the literary ones, the Maltese output for the news domain was significantly better than that for the literary texts. The translations generated in the technical domains contained fewer errors than for literature and news, but were still of inferior quality with respect to their English counterparts.

If the whole corpus is considered in its entirety, the percentages of errors appear as shown in Table 8 below:

Table 8: Total percentage of errors in the whole corpus

Google T. EN	ChatGPT EN	Google T. MT	ChatGPT MT
1.33	0.46	3.53	3.46

The totals show that error frequency for Google Translate was 2.65 times higher in Maltese than in English, and 7.5 times higher in the case of ChatGPT.

5.6 Considerations on the findings

The above data was collected from a 6000-word corpus of 20 texts from 4 domains. The 300-word limit set for each text was deemed representative of the whole text for the purposes of the study. However, these criteria constitute a limitation in some ways. Firstly, the texts chosen cannot be considered representative of the entirety of their respective domain. In the domains of literature and news, the writer’s style may affect the performance of the tools, while in financial and medical text a more consistent register is frequently used. Moreover, the subject matter of the text itself has an impact on the tools’ performance even within the same domain, especially in LRLs, where the availability of specific terminology or lack thereof strongly affects the quality of a target text. For example, in a particular LRL, a terminology may be available for the domain of diabetes but may be lacking in the case of orthopaedics, possibly due to the efforts by local associations, awareness campaigns or ongoing projects concerning diabetes and its treatment to use Maltese in their outreach. In such cases, the terminology would constitute available data that could be used for training the tools, thus impacting positively on the raw output. On the contrary, a lack of data would have a negative effect on translation quality even if the source text belongs to the medical domain. Such issues exist in the case of Maltese, where the pace in the development of terminology across domains and subdomains is inconsistent, thus leaving pockets of unavailable data that have an impact on translation driven by technology.

The quality issues discussed above have implications both on a professional and a pedagogical level. The higher number of errors to correct in LRL raw output requires more time and effort from post-editors, especially when the translation follows the source text’s syntactic structures too closely and the domain lacks established terminology. Moreover, there is a higher possibility of missing some less-evident errors that need correcting, which could impact the overall quality of the target text. On their part, translation trainers must be fully aware of the risks taken by students when they rely too heavily on automated translation tools, especially when LRLs are involved. A general tendency to overestimate the capabilities of AI is particularly evident in the domain of translation, and translation students themselves are not exempt from this inclination. Students must be informed of the specific issues faced by

LRLs and must be trained accordingly to develop post-editing skills that make them less inclined to be complacent towards the raw output provided by the tools.

6. Conclusion

The findings provide the first research-based evidence of the relatively low performance on Google Translate and ChatGPT in the translation of literary, news, financial and medical texts into Maltese. That both tools perform better with HRLs than LRLs has been shown by various studies (e.g. Goyle et al. 2023), but the present study gives figures that show the extent of the difference in translation quality between a HRL and Maltese, which must benefit fully from the resources available to the other official languages of the European Union. Unfortunately, such resources will continue to provide lesser quality as long as the training datasets available remain limited. The provision of data requires both financial and human resources that should be provided at a national level. A comprehensive national language policy, which to date has not yet been devised, should include measures to gather, manage and provide as many Maltese language resources as possible to improve the training of LLMs, thus increasing the quality of translation output.

Conflict of Interest Statement
The author declares no conflicts of interest.

About the Author
Sergio Portelli is a Professor at the Department of Translation, Terminology and Interpreting Studies (Faculty of Arts) at the University of Malta. He teaches Translation Technology, Terminology, Intercultural Communication and Translation Practice from Italian to English and Maltese. He is also a literary translator and a former translator of the European Parliament. His research interests also include Comparative Literature and Cultural Studies.

Works Cited

All Top Everything. (2025, October 9). Top 10 most used AI chatbots in the world (2025). All Top Everything. https://www.alltopeverything.com/most-used-ai-chatbots/ (last access 20 February 2026).

Ethnologue. (2025). Maltese (mlt). SIL Global. Retrieved June 2, 2026, from https://www.ethnologue.com/language/mlt/ (last access 10 November 2025).

Maltese Language Resource Server. (n.d.). MLRS corpora. University of Malta. Retrieved June 2, 2026, from https://mlrs.research.um.edu.mt/index.php?page=corpora (last access 4 March 2026).

Al-Romany, Tahseen and Maryam Jawad Kadhim. 2024. 'Artificial Intelligence Impact on Human Translation: Legal Texts as a Case Study'. *International Journal of Linguistics, Literature and Translation* 7: 89-95. https://doi.org/10.32996/ijllt.2024.7.5.11

Bhattacharja, Joydeep. 2026. 'Google Translate Statistics & Trends'. Retrieved from https://seosandwitch.com/google-translate-statistics/#rtoc-1 (last access 20 February 2026).

Cieri, Christopher, Mike Maxwell, Stephanie Strassel, and Jennifer Tracey. 2016. 'Selection Criteria for Low Resource Language Programs'. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*. Portorož, Slovenia: European Language Resources Association (ELRA), p. 4543–4549. Retrieved from https://aclanthology.org/L16-1720/

Dranch, Konstantin. n.d. 'Tools for Data Labelling in Machine translation Evaluations'. Retrieved from https://custom.mt/tools-for-data-labelling-in-machine-translation-evaluations/

Eszenyi, Réka, Klaudia Bednárová-Gibová and Edina Robin. 2023. 'Artificial Intelligence, Machine Translation & Cyborg Translators: A Clash of Utopian and Dystopian Visions'. *Orbis Linguarum* 21.2: 102-113. https://doi.org/10.37708/ezs.swu.bg.v21i2.13

Goyle, Vakul, Parvathy Krishnaswamy, Kannan Girija Ravikumar, Utsa Chattopadhyay and Kartikay Goyle. 2023. 'Neural Machine Translation for Low Resource Languages'. https://doi.org/10.1145/3567592

Hendy, Amr, Mohamed Abdelrehim, Amr Sharaf, Vikas Raunak, Mohamed Gabr, Hitokazu Matsushita, Young Jin Kim, Mohamed Afify and Hany Awadalla. 2023. 'How Good Are GPT Models at Machine Translation? A Comprehensive Evaluation'. https://doi.org/10.48550/arXiv.2302.09210

Il-Kunsill Nazzjonali tal-Ilsien Malti [National Council for the Maltese Language]. 2021. *The State of the Maltese Language: National Survey 2021*. https://kunsilltalmalti.gov.mt/wp-content/uploads/2023/03/rizultati-stharrig2021_MT-EN.pdf (last access 10 November 2025).

Karabayeva, Irina S. and Anna N. Kalizhanova. 2024. 'Evaluating machine translation of literature through rhetorical analysis'. *Journal of Translation and Language Studies* 5(1): 1-9. https://doi.org/10.48185/jtls.v5i1.962

Kemp, David. 2025. 'Digital 2026: Malta'. https://datareportal.com/reports/digital-2026-malta (last access 11 March 2026).

Mohamed, Yasir A., Akbar Kannan, Mohamed Bashir, Abdul Hakim Mohamed, Mousab A. E. Adiel and Muawia A. Elsadig. 2024. 'The Impact of Artificial Intelligence on Language Translation: A Review', *IEEE Access* 12: 25553-25579. https://doi.org/10.1109/ACCESS.2024.3366802

Moneus, Ahmed Mohammed and Yousef Sahari. 2024. 'Artificial intelligence and human translation: A contrastive study based on legal texts'. *Heliyon* 10: 1-14. https://doi.org/10.1016/j.heliyon.2024.e28106

National Statistics Office. 2024. *Malta Skills Survey 2022*. Malta: National Statistics Office, 129.

Portelli, Sergio. 2025. 'Terminology Development for Lesser-Used Languages in Bilingual Contexts: The Maltese Case'. In R. Resi and F. Steurs (eds). *Handbook of Terminology. Terminology Planning in Europe*. Amsterdam-Philadelphia: Benjamins, 653-669. https://doi.org/10.1075/hot.4.ter22

Shahmerdanova, Roya. 2025. 'Artificial Intelligence in Translation: Challenges and Opportunities'. *Acta Globalis Humanitatis et Linguarum* 2(1): 62-70. https://doi.org/10.69760/aghel.02500108

Shormani, Mohammed Q. 2025. 'Artificial intelligence contribution to translation industry: looking back and forward'. https://doi.org/10.1007/s44163-025-00487-3

Smith, Brad. 2025. 'Unlocking data to advance European commerce and culture'. https://blogs.microsoft.com/on-the-issues/2025/07/20/eudigitalunlock/