Cultural Transcreation in Asian Languages with Prompt-Based LLMs

Helena Wu[1,4], **Beatriz Silva**[1], **Vera Cabarrão**[1], **Helena Moniz**[2,3]
[1]Unbabel, Lisbon, Portugal
[2]University of Lisbon, Portugal
[3]CLUL, Lisbon, Portugal
[4]INESC-ID, Lisbon, Portugal
helenawu@edu.ulisboa,pt, {beatriz.silva, vera.cabarrao}@unbabel.com,
helena.moniz@edu.ulisboa.pt

Abstract

This research explores Cultural Transcreation (CT) for East Asian languages, focusing primarily on Mandarin Chinese (ZH) and the customer service (CS) market.[1] We combined Large Language Models (LLMs) with prompt engineering to develop a CT product that, aligned with the Augmented Translation concept, enhances multilingual CS communication, enables professionals to engage with their target audience effortlessly, and improves overall service quality. Through a series of preparatory steps, including guideline establishment, benchmark validation, iterative prompt refinement, and LLM testing, we integrated the CT product into the CS platform, assessed its performance, and refined prompts based on a pilot feedback. The results highlight its success in empowering CS agents, regardless of linguistic or cultural expertise, to bridge effective communication gaps through AI-assisted cultural rephrasing, thus achieving its market launch. Beyond CS, the study extends the concept of transcreation and prompt-based LLM applications to other fields, discussing its performance in the language conversion of website content and advertising.

1 Introduction

Transcreation, also known as creative translation, is a language conversion approach in which discussions remain relatively sparse and primarily emphasise manual approaches, with a focus on its application in fields such as advertising, marketing and literary translation (Díaz-Millón and Olvera-Lobo, 2023). The present research aims to expand this concept beyond human transcreation by developing an automatic transcreation product that incorporates cultural awareness through the adoption of prompt engineering and Large Language Models (LLMs), with a primary emphasis on Mandarin Chinese (ZH) — both Simplified (zh-CN-Hans) and Traditional (zh-TW-Hant) — along with Japanese (JA) and Korean (KO).

This research was carried out at Unbabel, which provides translation services widely used in industries such as customer support (CS) and e-commerce. The Cultural Transcreation (CT) product developed in this study is integrated into Unbabel's machine translation (MT) workflow as a pre-translation step, creating a new automated CT pipeline, as shown in Figure 1.

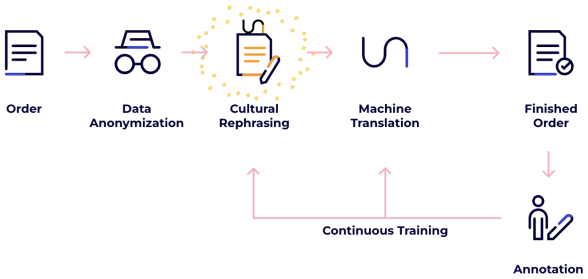

Figure 1: Automated CT Product Pipeline

The product is implemented on a CS platform, and is currently only available for use in the CS sector. The tickets (e-mails) produced by CS agents are first anonymised, and then undergo cultural adaptation within the source language to align with the target culture. Only after this cultural alignment are the tickets processed by the MT systems.

Notably, the principle of this product is aligned with the concept of Augmented Translation, which builds on Douglas Engelbart's (1962) vision of leveraging computational tools to enhance, rather than replace, human capabilities. In this context, our CT tool aims to equip CS agents — who communicate in the source language but lack cultural awareness of the target audience — with enhanced capabilities to effectively cross cultural and linguistic boundaries. Augmented Translation, as

[1]Part of this work was previously published in an international peer-reviewed venue, namely EAMT 2024. For the sake of clarity and ethical transparency, the following publication is integrated into the present work: Silva et al. (2024)

Proceedings of Machine Translation Summit XX Volume 2, pages 42–51, June 23-27, 2025

introduced by CSA Research and building upon Engelbart's integration of human expertise with AI-driven processes, represents a paradigm shift in the translation industry. It combines human expertise with advanced language technologies to enhance communication. Unlike traditional MT workflows, which rely solely on automation, this approach reinforces the role of human workers by integrating AI with human-in-the-loop processes. Rooted in Engelbart's (1962) vision of technology augmenting human capabilities, studies such as Lommel (2018) and O'Brien (2024) have further developed the concept, emphasising the necessity of a human-centered approach. Ultimately, our CT solution minimises misunderstandings caused by a single MT and the cultural differences that could lead to conflicts. It ensures that CS adheres more closely to the cultural conventions of the target language.

2 Literature Review

Transcreation, a portmanteau of "Translation" and "Creation", combines translation with creative adaptations to preserve the original message's essence while aligning it with specific target norms. This approach achieves a level of adaptation beyond what translation offers by addressing both creative and target-specific aspects. In the context of transcreation, the approach that focuses on cultural awareness can be called Cultural Transcreation, which is the central focus of this research.

In essence, transcreation bridges the gap between maintaining the source intention and addressing the cultural and audience-specific nuances. As Gaballo (2012) defines it: "Transcreation is an intra-/interlingual re-interpretation of the original work suited to the readers/audience of the target language, which requires the translator to come up with new conceptual, linguistic, and cultural constructs to make up for the lack (or inadequacy) of existing ones. It can be looked at as a strategy to overcome the limits of 'untranslatability'".

Along this line, we can develop the perspective that translation, localisation, and transcreation represent distinct yet overlapping approaches to language conversion, varying in their degree of adaptation and cultural sensitivity. While translation primarily emphasises on linguistic accuracy and fidelity to the source text, localisation adapts content to fit specific cultural and technical contexts, incorporating elements such as date formats and currencies (Pym, 2017, pp. 119, 131-132). Transcreation, by contrast, is a highly creative process that prioritises cultural adaptation, re-interpreting content to align with audience expectations while preserving the original intent. Though scholars such as Szasz and Olt (2018) and Rike (2013) clearly differentiate transcreation from other approaches, some (Mangiron and O'Hagan, 2006; Munday et al., 2022) consider it a subset of localisation.

Then, focusing on the perspective of distinguishing transcreation from the other two concepts, some theoretical frameworks, such as Juliane House's functionalism and Lawrence Venuti's critique, further elucidate these distinctions. House (1997, 2014, 2015) defines translation as a process to achieve functional equivalence between the source and the target, distinguishing between "overt" translations, which retain source-cultural elements, and "covert" translations, which adapt more seamlessly to the target culture. This functionalist perspective highlights that translation, though adaptive, primarily preserves the original content's intent and scope without significant creative re-interpretation. Localisation, aligning with skopos theory (Vermeer, 1978), extends beyond translation by embedding content within cultural, linguistic, and technical norms, ensuring usability without fundamentally altering the original's communicative intent. Transcreation, on the other hand, represents the most adaptive approach, focusing on emotional and cultural resonance over literal fidelity. Venuti's (2017) domestication theory is particularly relevant here, as transcreation prioritises audience engagement by reshaping content to align with target cultural expectations. From a functionalist perspective, transcreation aligns even more closely with skopos theory, as it prioritises intended impact, often requiring significant creative restructure.

In sum, distinguishing transcreation from the other two requires recognising their varying degrees of adaptation. While all three involve cross-cultural transformation, transcreation occupies the most dynamic end of this spectrum, ensuring content is not only translated but strategically reshaped to maximise cultural and emotional resonance.

In this way, transcreation, particularly CT, challenges the traditional boundaries of translation industry by addressing the need for cultural adaptation. The concepts of "high-context" and "low-context" cultures, introduced by Hall (1976), are vital for understanding cultural adaptation in language conversion. Low-context cultures, such as English-speaking countries, favour more ex-

plicit and direct communication. In contrast, languages of high-context cultures, such as Chinese and Japanese, rely heavily on implicit communication, shared assumptions, and contextual understanding. Effective translation between these frameworks requires more than linguistic accuracy, it demands cultural adaptation to ensure resonance with the target audience. Nida and Taber's (1969) concept of "dynamic equivalence" emphasises the impact over direct translation, particularly for high-context audiences. Katan (2015) notes that low-context information often requires adjustment to accommodate cultural subtleties, highlighting the critical role of CT in avoiding miscommunication.

The emergence and development of CT introduces several qualitative improvements compared to traditional translation methods. It enables creative adaptation that maintains the original message's essence while aligning with the cultural expectations of target audiences by prioritising key features of high-context cultures. Moreover, it mitigates the risks associated with hyper-literal translation, reducing potential misunderstandings and cultural misalignment. Beyond its academic value, transcreation holds significant economic potential for international markets. As noted by Carreira (2023), the rising demand for CT stems from non-language service provider companies addressing global communication challenges. It facilitates cross-cultural understanding, enhancing brand perception and customer engagement. Culturally sensitive communication strengthens customer relationships, ensuring content is engaging, relatable, and aligned with audience expectations. Moreover, integrating LLMs into transcreation workflows can amplify these benefits, enhancing cost and time efficiency while reducing reliance on manual transcreation, making it an economically viable solution for companies in global markets.

3 Methodology

This section outlines the systematic approach undertaken to develop a transcreative MT pipeline that incorporates cultural awareness into the translation workflow by leveraging prompt engineering and LLMs, and evaluates the effectiveness of prompt-based LLM transcreation. The research transitions from the theoretical aspects of transcreation to its practical implementation, addressing the current challenges in machine-generated cultural transcreation. After constructing guidelines and prompts tailored to the specific sector, these concepts will be integrated into the innovative product and applied in real-world scenarios to collect authentic data for evaluation and iterative optimisation of its market performance.

The core investigation and development area of this product is the CS sector. This field was chosen not only for the accessibility of relevant data but also because it offers pronounced cultural differences in communication between English-speaking CS agents and their international target audience. For this research, zh-CN-Hans, zh-CN-Hant, JA, and KO were identified as target languages, focusing specifically on Mandarin Chinese in this paper.

Considering English (EN) as the source working language, Mandarin Chinese as the target working language, and cultural transcreation as the research theme, we adopt a multi-stage experimental approach to achieve the study's objectives, where the outcomes of earlier experiments inform and support developing subsequent ones, creating an interconnected process. It is also worth mentioning that, in addition to all other responsibilities, the evaluation processes across all stages were also conducted manually and solely by the authors.

The first aim is establishing culturally aware guidelines based on shared assumptions within the target audience, manual observations of communication features, and real-world translation analysis. In addition, all actions were performed solely by the authors. Based on these guidelines, an initial version of the prompts was created to serve as the foundation for the subsequent experimental phases. The guidelines serve as the foundation for **benchmark analysis of MT cultural transcreation samples in the CS domain**, which assesses the effectiveness of initial prompts and LLM-generated outputs in the CS domain. By conducting a cultural assessment of automatic transcreation benchmarks, this stage improves prompt engineering strategies, which lead to establishing a formal Version 1.0 of the prompts, and inform subsequent experiments.

Then, the primary and central goal of this study — exploration and evaluation of the performance of the CT product generated by prompt-based LLM — will be achieved by conducting a **CT Clients Pilot**. This stage involves testing Version 1.0 of the prompts and refining it into Version 2.0 through seamlessly continuing data generation, real-time performance monitoring and adaptation.

Beyond CS, the study extends its last goal to assess the potential of prompt-generated LLM tran-

screation in other domains. By stepping beyond the primary focus on CS and cultural awareness, the exploration in **Website Content Generation and Advertising** aims to evaluate the adaptability and performance of automatic transcreation in distinct fields, contents, and text types, broadening the scope of its application and identifying future opportunities for research and development.

3.1 Basic Guidelines, Prompts, and LLM

We established a series of guidelines to implement the concept of cultural awareness, specifically for East Asian cultures. Since the initial target market is the CS sector, the guidelines were tailored to align with the cultural communication norms of this field. These guidelines included template examples, serving as strategic foundations to support the subsequent construction of prompts and experiments, enabling the LLM to generate culturally appropriate rephrasing more effectively.

The Chinese guidelines consist of three sections. The first section outlines specific communication methods and linguistic expressions to follow or avoid, with 16 suggestions for avoidance and 10 for compliance, comprising aspects such as text format, politeness, and emotional outputs.

The second section offers a more in-depth framework for CT in the CS field, covering key aspects such as e-mail formats, appropriate greetings and closings, and other practical expressions.

The final section presents real-world examples of manual cultural awareness annotations to further contextualise the guidelines. Native speakers of the target languages conducted culturally adapted rephrasing of actual CS e-mails, which were then used as templates for developing automatic transcreation. Furthermore, this section also includes additional resources, such as fictional CS e-mails designed as extreme case references for CT, as well as translation and rephrasing examples from chat interactions, offering a comprehensive foundation for refining the transcreation process.

Based on the established guidelines, the initial version of the prompts for e-mail rephrasing in the CS domain was successfully developed. These prompts were tailored for ZH, JA, and KO, reflecting each language's unique cultural requirements and were not shared across languages. Additionally, through iterative temperature tests, the optimal LLM temperature was identified as 0.7, adopting GPT-4 as the LLM. Notably, higher values tend to generate overly creative and unstable outputs, deviating from the prompts and becoming inconsistent with predefined expectations. In contrast, lower temperatures result in excessive rigidity, causing the LLM to either disregard prompts or fail to achieve the expected level of creativity.

3.2 Cultural Validation of MT Benchmarking

After establishing basic guidelines, prompts, and LLM settings, an MT benchmarking analysis was conducted to evaluate the initial version of the prompts. This step aimed to formalise the first production version of the prompts and assess the quality of culturally adapted transcreations generated by LLMs using the initial prompt.

The analysis is based on 21 MT benchmarking samples consisting of tickets provided by Unbabel clients across seven distinct industries, including social media platforms, food manufacturing, tourism, software development, electronic products, and gaming. Each sample includes both the original English text (Original EN), consisting of 166 segments, and the rewritten English text (Rephrased EN), comprising 174 segments. The Original EN represents the initial messages written by CS agents, while the Rephrased EN is produced by inputting the Original EN into an LLM, which rephrases the text to create a culturally sensitive and target-aligned EN version. This intermediary step ensures that the subsequent language translation process becomes more concise, clear, and tailored to the communication preferences of the target language audience. Moreover, this intermediate rephrasing step provides additional benefits beyond cultural alignment, strengthening the overall transcreation process in the CS domain. These aspects will be further elaborated in the next section, which focuses on the core client pilot experiment.

3.3 Cultural Transcreation Product Pilot

Following a brief benchmarking validation and the establishment of the first official prompts version for Asian languages, a three-month pilot was launched to gather authentic and real-time data, monitor product performance, and collect feedback for optimisation. The Cultural Transcreation concept was transformed as a cultural rephrasing feature, integrated into the CS platform via a "Rephrase" widget. This feature allows CS agents to click the button to access the CT service, which generates a rephrased EN version of the original text. The new version retains readability for CS agents while aligning with the target recipients'

cultural context. They can then decide whether to edit the rephrased text before sending it to the MT system. Upon sending, the recipient receives the final version that is more in line with their culture and communication style.

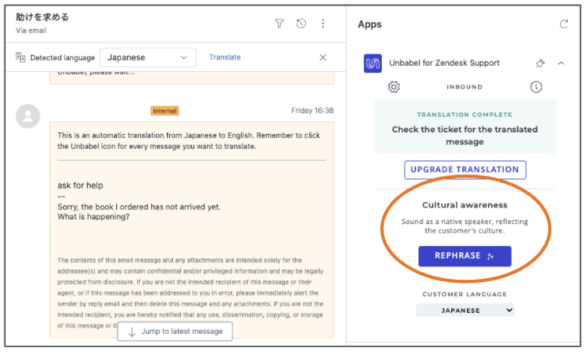

Figure 2: Interface of the CT Product Implemented in the CS Platform

In addition, the intermediate rephrasing step in the source language ensures greater accuracy and cultural relevance in subsequent translations while maintaining transparency of information and editability for the CS agents of the source language.

Figure 3: Pipeline of the CT product implemented in the CS Platform

This Live Pilot involved the CS teams of three Unbabel clients, selected for their high CS demands and frequent interactions with the international target groups. These companies represent three distinct industry sectors, providing a diverse dataset of email communications that reflect varied CS contexts and interactions. Specifically, **Client A**, **B**, and **C** belong to the Tourism, Internet and Electronics industries, respectively.

GPT-4 was initially selected as the LLM for the pilot, using a Few-Shot Prompting approach. The three-month pilot was divided into two phases. Phase I started on 26th January 2024. In the last week of this phase, we optimised the prompts based on collected feedback and re-evaluated alternative LLMs for rephrasing. Prompt Version 2.0 and the selected optimal model were deployed on 14 March, marking the start of Phase II, which continued until the pilot's official end on 29th April.

In terms of CT quality evaluation, the quality and performance of rephrased texts was assessed by the authors of the work through three key dimensions: (1) overall ticket quality, (2) errors caused by cultural rephrasing, and (3) the types and severity of rephrasing errors.

Firstly, the overall quality of the rephrased e-mails is classified into four categories: **Good** – successfully culturally transcreated e-mails without any error caused by LLM-generated rephrasing; **Minor issues** – contain minor errors with limited impact on clarity, yet culturally appropriate; **Major issues** – the rephrased text contains one or more notable errors, some of which affecting meaning or interpretation; **No change** – absence of meaningful cultural adaptation and lacks alignment with the communication norms of the target audience.

Secondly, we annotate errors from e-mails labelled "Minor Issues" or "Major Issues" in the previous step. It facilitates systematic data collection to inform the third dimension. Identified rephrasing errors are categorised into 11 types: **Greeting**, **Closing**, **Register**, **Added Information**, **Removed Information**, **Changed Information**, **Glossary Term**, **Emotion, Inconsistency**, **Grammar Issues**, and **Unidiomatic Expression**.

Then, errors are further assessed based on their severity: **Minor** (minimal impact on comprehension); **Major** (more pronounced issues that may alter meaning or reduce clarity); and **Critical** (causing misinterpretation or communication failure).

This structured evaluation framework ensures a consistent and rigorous assessment of LLM-generated rephrasings, enabling further optimisation of the transcreation process.

3.4 Discovering Transcreation on Website Content Generation

A preliminary investigation was conducted into another potential area for application: website content transcreation. This exploration aimed to evaluate whether transcreation, extending beyond cultural aspects to encompass broader bilingual transcreation, could be effectively applied to this field. As a side note, both this experiment and the subsequent supplementary advertising test were conducted exclusively in Simplified Chinese.

The source data for this exploration was provided by a cloud solution provider. For the trial dataset, based on more than 4,000 segments provided by the platform, we randomly selected 20 translation segments with human translation (HT) from each of seven different websites using the platform for content translation, with HT serving

as the gold standard. This resulted in a dataset of 140 segments. Unlike the CS sector tests, this trial simplified the transcreation process by instructing the LLM to directly transcreate the source text into the target language, bypassing the intermediate EN rephrasing step, as shown in Figure 6.

Figure 4: Pipeline of Transcreation Process Adopted in Website Content Generation

In terms of the seven websites from diverse industries, each presents distinct content features: Website A is an e-commerce platform offering contact lenses and selfie phones. Website B, a GPS tracker retailer, consists of marketing texts and product descriptions. Website C, a mattress store, provides product specifications and usage guidelines. Website D supports bio-pharmaceutical partnerships. Website E is a men's fashion retailer comprising product information and navigation-related content. Website F offers data science training content, and its dataset relies primarily on HT, with only two segments adopting MT.

Based on an initial review of the translation data, we developed a set of prompts for direct source-to-target (en > zh-CN) transcreation of multi-domain website content, employing the Few-Shot Prompting technique. GPT-4o, OpenAI's latest LLM at the time, was selected as the designated model for the transcreation tasks. This decision followed brief comparative tests with various LLMs, where GPT-4o displayed superior performance and optimal compatibility with the prompt. After testing different configurations, the model's temperature was set to 1.0 to achieve the desired balance between creativity and stability in the outputs.

As noted before, human translation for each data segment was defined as the gold standard for evaluating the quality of transcreated outputs generated by prompt-based LLM. This optimal reference enabled a comparative analysis, with quality evaluation annotations conducted by the author. The ratings were divided into five distinct categories, ranging from high to low, each accompanied by abbreviations for use in graphical representations in the Results and Discussion chapter:

1. Transcreation quality superior to human translation (**Tc > HT**)
2. Transcreation quality equal to human translation (**Tc = HT**)
3. Transcreation quality lower than human translation but higher than Machine Translation (**HT > Tc > MT**)
4. Transcreation quality equal to Machine Translation (**Tc = MT**)
5. Transcreation quality lower than Machine Translation (**Tc < MT**)

3.4.1 Additional Test in the Advertising Field

After completing the investigation on website content, a supplementary test was conducted. 20 segments in EN with creative potential of publicly available marketing slogans and advertising phrases were randomly selected. The aim was to evaluate whether the developed prompt, in conjunction with the same LLM (GPT-4o) and its configuration, could effectively perform automatic transcreation in the advertising domain, a field known for its complexity and creative demands, typically making it more suitable for human transcreation.

Additionally, given that some LLMs may inherently exhibit creative rewriting or translation skills that go beyond direct translation, it is crucial to attribute all transcreation results in this supplementary test solely to the prompt engineering developed in this research, rather than the LLMs' inherent abilities. To facilitate a clear comparison and highlight the impact of the developed prompt, the study also includes a baseline translation of the selected texts from EN to ZH without the use of prompts. Each translation/transcreation will include a hyper-literal EN back-translation of the Chinese output for additional clarity. However, it is important to note that these back-translations cannot fully convey the nuances and linguistic subtleties inherent in the Chinese text, as many differences in Chinese expression and phrasing are not directly translatable. This limitation should be considered when interpreting the results in the subsequent chapter.

4 Results and Discussion

4.1 MT Benchmarking for CT

As outlined in the Methodology chapter, this section examines 21 MT benchmarking samples from Unbabel clients across seven distinct industries. The samples were first culturally rephrased into EN to align with the cultural nuances of the target

language, then translated into ZH, incorporating these cultural adaptations.

The analysis begins by categorising the types of modifications observed in the culturally rephrased text segments. Following this, a detailed examination will be conducted for each cultural rephrasing type, focusing on identifying cultural adjustments that remain absent or require further refinement, as highlighted through observations of the 21 samples. Note that these samples include 166 segments of Original EN and 174 segments of Rephrased EN.

The rephrased EN group demonstrated extensive modifications, including the addition of contextually relevant information to address gaps in the Original EN group. However, not all adjustments enhanced cultural adaptations. For statistical analysis, only culturally valid modifications contributing to cultural optimisation are considered.

Across all 21 tickets, 103 culturally rephrased segments were identified, representing approximately 59.20% of the total segments in the Rephrased EN tickets. As shown in Table 1, these 103 segments were categorised into four primary types of valuable rephrasing. It is essential to note that multiple rephrased segments may appear within a single ticket, and individual segments may exhibit several rephrasing types. Therefore, the Grand Total in the table reflects the total number of culturally adapted segments across all tickets, rather than the sum of rephrasing occurrences.

Main Category	Subcategory	No. of Segments	No. of Tickets
Politeness Adjustment	Courtesy Words	9	7
	Requests/Offers/Invitations	5	5
	Salutations and Valedictions	25	21
	Emotional Outputs	12	9
	Euphemism	0	0
	Total	**51**	**21**
Paraphrasing	Total	**56**	**19**
Grammatical Person Adjustment	Total	**3**	**3**
Functional Equivalence	Total	**4**	**3**
Grand Total		**103**	**21**

Table 1: Statistical Data of Rephrasing Types in Rephrased EN

Among the identified rephrasing types, **Politeness adjustment** emerged as a prominent cultural adaptation, with 51 segments across 21 tickets undergoing this modification.

While Politeness Adjustment was prevalent, **Paraphrasing** was the most common rephrasing type, observed in 56 segments across 19 tickets. This involved rewriting sentences to enhance clarity and cultural appropriateness. For example: Original EN – “Due to a system limitation, we can only reply in Chinese Simplified”; Rephrased EN – “Unfortunately, due to a system limitation, our reply will only be available in Chinese Simplified”.

Another category, **Grammatical person adjustment**, was found in 3 segments across 3 tickets. These adjustments involved switching from singular to plural pronouns (e.g., “we” instead of “I”) to align with CS conventions in ZH.

Functional equivalence accounted for 4 segments in 3 tickets. This category replaced idiomatic phrases in the source language with culturally appropriate equivalents in the target language. For instance: Original EN – "Consumer care channel”; Rephrased EN: “Customer service channel”.

Furthermore, the analysis revealed 6 segments across 6 source tickets requiring cultural adaptation but left unaddressed: 1. **Functional equivalence** (4 segments); 2. **Emotional outputs** (1 segment); 3. **Paraphrasing** (1 segment).

Additionally, the structural analysis of the rephrased samples culminated in the creation of a mind map categorising all observed types of CT rephrasing. This framework offers a comprehensive overview of rephrasing types and their applications, serving as a foundation for future research and practical advancements in CT.

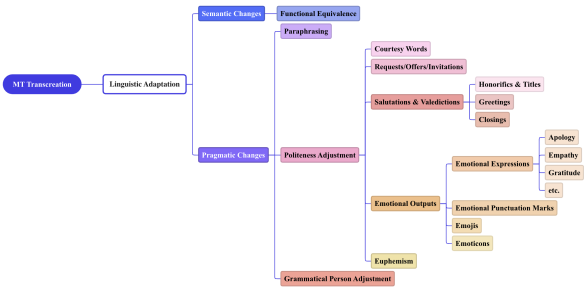

Figure 5: Linguistic-Based Classification of Transcreation Rephrases

4.2 Live Pilot Programme

As presented in the Methodology chapter, this pilot was divided into two phases. Phase I utilised the version 1.0 of the prompts, while Phase II introduced the optimised Prompt 2.0, developed from feedback in Phase I. The statistical data below summarises the culturally transcreated tickets produced during the three-month pilot. A total of 352 e-mails were reformulated using the CT product from EN to ZH, with 60.80% of the data collected during the first phase and 39.20% during the second phase. In addition, it should be noted that the decrease in data collection during Phase II was attributed to external factors such as the holiday season among CS agents and internal organisational restructuring.

	Target Language	No. of Segments	No. of Tickets	Total
Client A	zh-CN-Hans	3	9	12
	zh-TW-Hant	11	21	32
	Total	14	30	**44**
Client B	zh-CN-Hans	96	35	131
	zh-TW-Hant	26	12	38
	Total	122	47	**169**
Client C	zh-CN-Hans	8	6	14
	zh-TW-Hant	70	55	125
	Total	78	61	**139**
Grand Total		214	138	**352**

Table 2: CT Pilot Data Statistics – Number of Tickets

Despite these challenges, the comparative results between the two phases were promising. Following the two phases of pilot, using the same LLM but different prompt versions, a clear comparison is shown in Figure 6. While Prompt 1.0 already demonstrated outstanding performance, with no e-mails categorised as “No changes” (lacking cultural adaptation), the optimised Prompt 2.0 further improved the quality of automated CT. The percentage of e-mails rated as performing perfectly (“Good”) increased from 40.65% to 57.97%, exceeding half of the total. Meanwhile, the proportion of e-mails with "Minor issues" and "Major issues" decreased significantly, from 42.06% to 32.61% and 17.29% to 9.42%, respectively. In other words, the percentage of e-mails achieving satisfactory quality ("Good" and "Minor issues") rose from 82.71% to 90.58%. These improvements underscore the success of the prompt optimisation process and highlight how well-designed prompts can significantly improve the performance of a product such as CT when used with consistent LLMs.

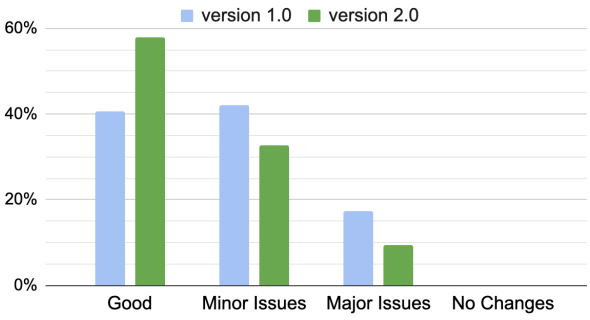

Figure 6: Rephrasing Quality of E-mails in Phase I & II

Regarding segments with rephrasing errors in the two phases, the results show a substantial reduction in errors from Prompt 1.0 to Prompt 2.0, particularly in categories such as *Greeting*, *Changed & Removed Information*, and *Unidiomatic Expression*. The error rate for *Greeting* decreased from 1 error per 23.78 tickets in Phase I to 1 error per 27.6 tickets in Phase II. Errors related to *Changed Information* dropped from 1 per 5.49 tickets to 1 per 34.5 tickets, while errors in *Removed Information* decreased from 1 per 15.29 tickets to 1 per 17.25 tickets. *Unidiomatic Expression* saw a significant decline, with errors reducing from 1 per 53.5 tickets to 1 per 138 tickets. Furthermore, errors in *Register*, *Added Information*, and *Emotion* were entirely eliminated in Phase II, improving from an initial rate of 1 error per 53.5 tickets, 214 tickets, and 53.5 tickets, respectively. Additionally, the number of "Major errors" saw a significant reduction, improving from an average of 1 error per 5.1 tickets in Phase I to 1 per 10.62 tickets in Phase II.

Then, by similarly calculating the average rephrasing error rate per ticket in both phases, the research reached the following result: in Phase I, the rephrasing error rate was 0.58 per ticket, which equates to 1 error for every 1.74 tickets. In Phase II, the rephrasing error rate decreased to 0.49 per ticket, meaning that an error occurred only once for every 2.06 tickets, while it is important to note that the probability of this being a “Major Error” is lower than in the previous phase.

In sum, the percentage of e-mails with optimal quality has increased substantially, while the occurrence of rephrasing errors has decreased. This not only an indicates the success of our prompt optimisation, but also illustrates how well-designed prompts can substantially improve the performance of products like CT when using the same LLM. Lastly, while this paper focuses on the results in ZH, similar improvements were observed in both JA and KO. Comparisons between the first and second versions of the prompts revealed notable progress and strong results across these languages.

4.3 Website Content Transcreation

The comparison of transcreated segments with the original website translations revealed that only 10 segments (7.14% of the total) exhibit quality between HT and MT, suggesting that the prompt-based transcreation approach introduced in this study outperforms standard MT. The remaining 92.86% (130 segments) were evenly distributed across the two highest levels. Of these, 65 segments (46.43%) exceeded the HT benchmark, indicating a quality level surpassing the defined gold standard. The other 65 segments (46.43%) matched HT quality. Notably, while $Tc > HT$ is ranked above $Tc = HT$ in quality evaluation, this does not imply that all segments can surpass HT. In some cases, $Tc = HT$ represents the highest attainable quality, as the

human-translated dataset already provides optimal translations. Therefore, segments matching HT further validate the effectiveness of the prompt-based LLM transcreation approach. These findings highlight that the prompt developed in this test, coupled with GPT-4o, has the potential to replace a substantial portion of HT in website content generation.

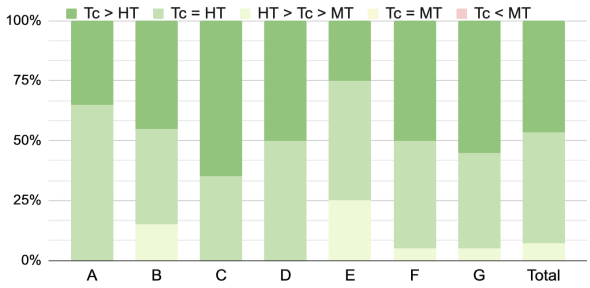

Figure 7: Quality of Transcreated Segments per Website

4.3.1 Transcreation in the Advertisement Industry

Owing to length constraints, Figure 9 shows only 3 of the 20 advertising segments of this test, where green-coloured segments highlight strong creative elements, red-coloured ones denote failed or awkward translations, and underlined sections indicate parts where the sentence is unnatural, lacks fluency, or does not align with the essential linguistic features of effective Chinese advertising. In the remaining unpresented slogans, prompt-based transcreation generally outperformed transcreation without prompts. Only 3 segments exhibited identical quality, as both represented optimal translation choices. This supplementary test clearly demonstrated the potential of integrating transcreation prompts with GPT-4o to achieve exceptional levels of creative adaptation and translation.

Brand	Original slogan	GPT-4o without prompt	GPT-4o with prompt
Esso	*Put a Tiger in Your Tank*	*在你的油箱里加一只老虎*	*为您的油箱加点猛虎之力*
		EN: "Add a tiger in your tank"	EN: "Give your fuel tank the power of tiger"
Tide	*Tide's in, dirt's out.*	*汰渍到，污渍跑。*	*汰渍到，污渍掉。*
		EN: "Tide arrives, stains run."	EN: "Tide arrives, stains wash out."
Dollar Shave Club	*Shave Time. Shave Money.*	*省时间，省钱。*	*省时省钱，刮须无忧。*
		EN: "Save time, save money."	EN: "Save time save money, shave without worry."

Figure 8: Advertising Slogans' Transcreation Outputs

In conclusion, the success of the website content prompt developed at this stage is highly significant, demonstrating its ability to effectively process diverse content and text types across multiple fields.

5 Conclusions and Future Work

By bridging linguistics and AI, this research has made significant strides, foremost among them being developing an automated CT system for the CS sector. To achieve this, culturally aware guidelines were established for three Asian languages, followed by the evaluation of the initial version of the prompts through their automated transcreation outputs using MT benchmark samples. These prompts were then further refined, alongside the creation of a linguistically based classification for categorising transcreation rephrasing. The CT pilot test confirmed that the continuous optimisation of prompt-based LLMs significantly enhanced the cultural transcreation quality. This showed the product's potential as a valuable AI-assisted tool for real-world applications, increasing the productivity and efficiency of CS agents in their communication tasks. Beyond cultural adaptation, this study also explored the feasibility of prompt-based automated transcreation for website content and advertising. As a result, we developed a successful multi-purpose prompt adaptable across industries, content types, and text genres. These findings lay a solid foundation for future advancements and product innovation, with the goal of expanding transcreation applications beyond CS to drive broader industry adoption.

At present, efforts are underway to integrate Unbabel's proprietary TowerLLM (Alves et al., 2024) into the CT product to replace third-party LLMs. After optimising Prompt 3.0 and training the internal model, TowerLLM was tested with 21 EN-ZH ticket samples to generate CT outputs. These outputs were compared with GPT-4o outputs for the same samples to identify the best culturally adapted versions. The results showed that TowerLLM produced the best transcreated versions for 17 of 21 samples, with the remaining 4 being a tie between the two models. This progress indicates a potential shift to TowerLLM as the product's primary model soon. Additionally, development is underway for an automated CT quality monitoring programme, with ongoing efforts to explore broader areas of transcreation to expand Unbabel's services.

From a broader AI industry perspective, we believe that the sensible and responsible integration of AI into language services fosters human progress, as pursued by the CT product developed in this study. This aligns with the Augmented Translation concept, which views AI as a collaborator

rather than a replacement for human translators, enhancing workflows and enabling creative problem-solving. In the CS sector, this means that CS agents are not replaced by automation but are instead empowered with AI-driven tools that support decision-making, refine translations, and adapt responses to align with cultural expectations. By leveraging this approach, the CT product not only facilitates the work of CS agents and enhances their efficiency, but also mitigates potential challenges arising from communication styles and cultural differences. Crucially, it adjusts tone, contextual appropriateness, and subtle linguistic distinctions vital for effective cross-cultural communication. To conclude, maximising AI's benefits requires a balanced strategy — leveraging its potential while maintaining human agency, ensuring AI enriches rather than disrupts professional and social progress.

Acknowledgments

This work was conducted as part of project no. 62 – "Center for Responsible AI", funded by European Funds under the "Recovery and Resilience Plan - Component 5: Agendas Mobilizadoras para a Inovação Empresarial", within the NextGenerationEU funding program. It received partial funding from FCT (Fundação para a Ciência e a Tecnologia) under project UIDB/50021/2020 (DOI:10.54499/UIDB/50021/2020). Additionally, the work was supported by the Centro de Linguística da Universidade de Lisboa (UID/00214).

References

Duarte M. Alves, José Pombal, Nuno M. Guerreiro, Pedro H. Martins, João Alves, Amin Farajian, Ben Peters, Ricardo Rei, Patrick Fernandes, Sweta Agrawal, Pierre Colombo, José G. C. de Souza, and André F. T. Martins. 2024. Tower: An open multilingual large language model for translation-related tasks. *Preprint*, arXiv:2402.17733.

Oliver Carreira. 2023. Surveying the economics of transcreation from the perspective of language professionals. *Across Languages and Cultures*, 24:127–144.

Mar Díaz-Millón and María Dolores Olvera-Lobo. 2023. Towards a definition of transcreation: a systematic literature review. *Perspectives*, 31(2):347–364.

Douglas C. Engelbart. 1962. Augmenting Human Intellect: A Conceptual Framework.

Viviana Gaballo. 2012. Exploring the boundaries of transcreation in specialized translation. *ESP Across Cultures*, 9:95–113.

Edward T. Hall. 1976. *Beyond Culture*. Anchor Books. Knopf Doubleday Publishing Group.

Juliane House. 1997. *Translation Quality Assessment: A Model Revisited*. Tübinger Beiträge zur Linguistik. G. Narr.

Juliane House. 2014. *Translation Quality Assessment: Past and Present*, pages 241–264.

Juliane House. 2015. *Translation: A Multidisciplinary Approach*. Palgrave Advances in Language and Linguistics. Palgrave Macmillan UK.

David Katan. 2015. Translation at the cross-roads: Time for the transcreational turn? *Perspectives*, 24:1–16.

Arle Lommel. 2018. Augmented translation: A new approach to combining human and machine capabilities. In *Proceedings of the 13th Conference of the Association for Machine Translation in the Americas (Volume 2: User Track)*, pages 5–12, Boston, MA. Association for Machine Translation in the Americas.

Carme Mangiron and Minako O'Hagan. 2006. Game localisation: Unleashing imagination with 'restricted' translation. *JOURNAL OF SPECIALISED TRANSLATION*, 6.

Jeremy Munday, Sara R. Pinto, and Jacob Blakesley. 2022. *Introducing Translation Studies: Theories and Applications*. Taylor & Francis.

Eugene A. Nida and Charles R. Taber. 1969. *The Theory and Practice of Translation*. Helps for translators. E. J. Brill.

Sharon O'Brien. 2024. Human-centered augmented translation: against antagonistic dualisms. *Perspectives*, 32(3):391–406.

Anthony Pym. 2017. *Exploring Translation Theories*. Online access with EBA: Taylor & Francis. Taylor & Francis.

Sissel M. Rike. 2013. Bilingual corporate websites-from translation to transcreation?

Beatriz Silva, Helena Wu, Yan Jingxuan, Vera Cabarrão, Helena Moniz, Sara Guerreiro de Sousa, João Almeida, Malene Sjørslev Søholm, Ana Farinha, and Paulo Dimas. 2024. Cultural transcreation with LLMs as a new product. In *Proceedings of the 25th Annual Conference of the European Association for Machine Translation (Volume 2)*, pages 57–58, Sheffield, UK. European Association for Machine Translation (EAMT).

Maria A. Szasz and Maria C. Olt. 2018. *Translation, Transcreation and Localization*. Editura U. T. Press.

Lawrence Venuti. 2017. *The Translator's Invisibility: A History of Translation*.

Hans J. Vermeer. 1978. Ein rahmen für eine allgemeine translationstheorie. *Lebende Sprachen*, 23(3).