

IMPROVING REAL-TIME LOW-RESOURCE LANGUAGE EUPHEMISM TRANSLATION WITH AGENT KNOWLEDGE BASE

ZiQing Zhong, XiaoMing Lu*, TianTian Gong

School of Interpreting and Translation Studies, Guangdong University of Foreign Studies, Guangzhou 510420, Guangdong, China.

**Corresponding Author: XiaoMing Lu*

Abstract: Translating euphemisms and politeness expressions between languages with distinct cultural norms presents significant challenges for automated translation systems, particularly for low-resource language (LRL) pairs. This study examines the effectiveness of integrating a custom-built knowledge base (KB) into large language models (LLMs) to enhance the translation of euphemistic expressions and politeness markers between Persian, Chinese, and English, with a focus on the Persian cultural schema of *taarof*. Two qualitative translation assessments were conducted: a Persian ceremonial speech translated into Chinese, and English audiovisual subtitles translated into Chinese and subsequently into Persian. Five translation versions were produced using LLMs with and without a domain-specific knowledge base, and evaluated by one Chinese lecturer and two native Persian speakers. Findings reveal that LLMs integrated with the knowledge base consistently outperform standalone models in terms of semantic accuracy and politeness transfer. However, divergent evaluations between evaluators highlight that politeness perception remains culturally contingent. This study demonstrates the value of knowledge base integration for LRL translation while underscoring the continued necessity of human's machine translation post editing efforts.

Keywords: *Taarof*; Low-resource languages; Knowledge base; Large language models

1 INTRODUCTION

Despite growing interest in cross-cultural pragmatics, few studies have examined Persian politeness, and even fewer have explored its implications for automated translation. This gap is significant given that translation is fundamentally intertwined with culture, particularly when rendering euphemisms and politeness markers - linguistic strategies employed to express deference, mitigate face-threatening acts (FTAs), and navigate social relationships. What constitutes an appropriate euphemism in one language may be perceived as inappropriate in another, and this challenge is amplified when translating between languages with fundamentally different politeness systems. Persian *taarof* - a cultural schema involving ritualized, often exaggerated, honorifics expressing humility and deference - offers a compelling case study, as it manifests in speech acts where what is said often diverges significantly from what is meant, creating translation challenges that extend beyond lexical equivalence to pragmatic and cultural dimensions. While large language models (LLMs) have improved automated translation for high-resource language pairs, low-resource pairs such as translations between Persian and Chinese lack comparable textual resources and unified evaluation standards. Furthermore, LLMs often struggle with culturally embedded pragmatic features, producing translations that may be semantically accurate but culturally inappropriate.

This study investigates how effectively current LLMs translate euphemisms and politeness expressions between Persian, Chinese, and English, and whether integrating a domain-specific knowledge base (KB) improves translation quality. Two qualitative assessments were conducted: a Persian ceremonial speech translated into Chinese, and English audiovisual subtitles translated into Chinese and subsequently into Persian. Five translation versions were generated using LLMs with and without a self-constructed KB containing peer-reviewed articles on politeness and euphemism conventions. Translations were evaluated by one Chinese lecturer specializing in Persian instruction and two native Persian speakers, enabling cross-cultural examination of translation appropriateness regarding politeness norms and euphemistic expressions. Beyond its theoretical contributions to translation studies and cross-cultural pragmatics, this research offers practical insights for translation practitioners, localization specialists, and AI developers seeking to enhance low-resource language (LRL) automated translation systems for culturally sensitive content. By demonstrating the potential of KB integration, this study also contributes to ongoing discussions about human-machine collaboration in translation workflows, suggesting that domain-specific resources of knowledge enhancement may serve as valuable preparatory tools before human post-editing.

2 OVERVIEW OF EUPHEMISM IN SPEECH ACTS

The term “euphemism” derives from Greek, generally understood as the use of gentle, friendly expressions to convey content that somehow might be sharp, abrupt, or otherwise uncomfortable to address directly. In previous research, linguists categorized euphemisms into positive and negative ones. Positive euphemisms elevate the semantic meaning of the original term, while negative euphemisms are typically employed to mitigate or eliminate the negative

connotations associated with taboo words. Either of them fulfills communicative and aesthetic functions. Among interpersonal interactions, euphemism is characterized by politeness and respect for others, emphasizing other's psychological needs, sometimes over ourselves' in the context of Pan-Asian culture. Conversely, certain deceptive euphemisms are commonly utilized in military, political, and economic discourse to obscure intentions and cover up failures or mistakes. Overall, the positive function of euphemism lies in facilitating harmonious interpersonal relationships, whereas its negative counterpart includes confusing right and wrong, beautifying disgraceful actions, and disguising their true nature [1].

Euphemisms, particularly those relating to taboos, are employed in a natural and widely used manner when certain expressions are avoided for reasons of propriety or social constraint. In ancient China, for example, it was forbidden to use syllables resembling the names of members of the imperial family. Euphemisms also function as expressions of respect and etiquette: instead of saying "your daughter", speakers may say "your precious one"; and rather than stating directly "she is pregnant", they may imply "she has joyful news", or "she has a joyful pulse". Similarly, it is considered improper to say "to go to the toilet" or "to urinate" in many contexts; the euphemistic "to relieve oneself" is preferred. As in other cultures such as Persian and English, Chinese has long exhibited a tendency to avoid words homophonous with "death". In contemporary Chinese, expressions such as "went to eternal rest", "has gone", "no longer alive", "left this world", or "went to see the ancestors" are used instead of the blunt term "died". Classical literature also abounds in euphemistic designations: a thief might be referred to via the four-character idiom literally meaning "gentleman on the roof beam", and a sex worker might be described as "a girl who lives in a building decorated with green lacquer". Comparable refined diction is found in modern Chinese, where drug addicts may be called "gentlemen with addiction". As Stanislav observes, the functions of hiding, masking, or veiling are highly frequent in Chinese euphemism, typically motivated by sympathy, optimism, and a desire to soften the emotional impact of unpleasant or socially undesirable realities [2].

Politeness, a fundamental feature of communication studied in sociolinguistics, pragmatics, and intercultural communication, also varies significantly across cultures. British politeness tends to emphasize respect for privacy, personal space, independence, and autonomy; in this context, offers of help may be perceived as patronizing, as they can imply doubt about the addressee's competence or self-sufficiency. In contrast, Persian politeness is closely tied to adherence to social etiquette, especially the strict politeness system known as *taarof*, which involves compliments, courtesy, and highly respectful forms of address.

Persian culture places a premium on maintaining solidarity, keeping close contact, and sustaining interdependence; offering help and openly sharing one's problems are understood as signs of goodwill rather than invasions of privacy, and respect for elders and higher-status individuals is strongly emphasized. These cultural values are reflected in family discourse patterns. In British families, communication between parents and children is relatively symmetrical, with indirect requests and expressions of thanks occurring in both directions. In Persian families, however, greater asymmetry is evident: children address parents indirectly and politely to signal respect and prompt obedience, whereas parents more often employ direct or imperative forms. Similarly, while British family members tend to thank one another with similar frequency and form, Persian children thank their parents more frequently and elaborately than parents thank their children [3].

Evaluations of (im)politeness are grounded in expectations shaped by individuals' cognitive dispositions and accumulated life experiences, which means that perceptions of politeness and impoliteness can differ markedly from person to person. These individual differences are further amplified by cultural factors, as even within a single cultural community, norms and expectations regarding (im)politeness and euphemistic usage are neither uniformly understood nor uniformly accepted [4].

3 CHARACTERISTICS OF POLITENESS EXPRESSIONS IN PERSIAN

Taarof functions as a cultural schema that enables speakers to perform face work and negotiate social relationships through speech acts including making requests, offering and accepting goods and services, and extending invitations. Violating this schema can be considered impolite depending on context and interpersonal dynamics [5]. Fixed politeness expressions are widely used in Persian social occasions and rituals, appearing in common interactions such as greetings, farewells, gratitude, and blessings, as well as specific interpersonal exchanges. For example, when offered tea, a guest may say "نکته درد دستت" (May your hand not ache) as "thank you", and the host responds with "درد سرت" (May your head not ache) as "you're welcome", with such phrases recurring throughout conversations to reinforce politeness.

Persian politeness expressions also demonstrate morphological and syntactic flexibility, notably distinguishing between singular (informal) and plural (formal/polite) forms of the pronoun *you*. In contrast, English uses only one second-person pronoun *you* regardless of formality or social context. Although Mandarin Chinese possesses both an informal second-person pronoun *nǐ* and a formal counterpart *nín* - developed in the 20th century to mirror Western deferential forms - the latter appears mainly in spoken discourse in northern China and is rarely used in casual online communication, though it remains common in professional or administrative contexts [6]. Translating FTAs between Persian and other languages is particularly challenging due to significant cultural variations in communicative behaviors. Requests, a major category of FTAs, threaten the hearer's negative face by impeding their freedom of action, and translators often err by overlooking politeness aspects and cultural differences in request formulation. To minimize face-threats, speakers employ indirectness through negative politeness and off-record strategies, using ambiguity to

address the hearer's desire for autonomy. Since more threatening acts require greater indirectness, and direct requests are perceived as impolite for intruding upon the addressee's territory (either physical or emotional one), cultural differences in request strategies frequently cause cross-cultural misunderstandings in both written and spoken translation. Recognizing these complexities is essential for applying appropriate translation strategies [7-8].

In Persian culture, the concept of over-politeness can be observed in *taarof*, which involves reciprocal exchanges of ritualized and often exaggerated honorifics expressing goodwill, flattery, and formalities. This reflects a moral order of prioritizing others' feelings, needs, and desires at the expense of one's own. This practice encompasses various interactional behaviors including real and ostensible invitations and offers, refusals, compliments and their responses, and showing reluctance to accept money [9]. However, evaluations of *taarof* vary from politeness to over-politeness depending on the participants and context. While some instances are considered perfectly appropriate, others are subject to self-criticism in Persian discourse for being overdone and unnecessary. For instance, offering food to strangers using “بفرمایید” (please help yourself) is a common *taarof* practice. However, this offer is ostensible rather than genuine - both parties understand it is merely a polite gesture, and the stranger is expected to refuse with thanks. The reason is that accepting food from a stranger would be considered socially awkward or inappropriate. While many Iranians view this as appropriate social etiquette, others criticize it as an empty ritual that lacks sincerity when extended to strangers [10].

4 KNOWLEDGE BASE INTEGRATION IN AUTOMATIC TRANSLATION SYSTEMS

LLMs encode vast amounts of factual knowledge in their parameters, but this knowledge is not organized for fine-grained, domain-specific reasoning or for detailed pragmatic analysis. By integrating an external KB, translation systems can access explicit, fine-grained information that is critical for disambiguating ambiguous terms and ensuring factual accuracy. Structured knowledge directly addresses translation challenges such as named-entity disambiguation and domain-specific terminology alignment, which are often unreliably handled by vanilla LLMs because their knowledge is implicit and non-transparent. For translation systems, such precision is crucial: incorrect relational knowledge, for example, mislinking “Apple” to the fruit rather than the company, can propagate errors throughout the output, whereas a reliable KB can function as a fact-checking layer that enforces consistency of entities and relations.

Integrating external KBs also helps mitigate well-known limitations of LLM-based translation, including outdated knowledge, hallucinations, and weak coverage of rare entities, particularly in LRLs. The impact of KB integration varies across different LLMs. Combining a strong KB with a high-performing model tends to yield the largest gains, especially for specialized LRL translation tasks. In this setting, the KB functions as a symbolic memory for the translation system, storing user- and domain-specific terminology, herein for example, culture-specific euphemisms that the LLM may not reliably encode. This allows systems to adapt to specialised LRL translation needs without retraining the underlying model [1,11]. This avoids the enormous cost of parameter adjustments within LLMs.

In this study, we contrasted vanilla LLM usage with KB-augmented usage by means of different prompting strategies. For the vanilla condition, we used a minimal prompt: “Please translate the following text into Persian (Chinese). The translation should be presented in a table: the first column is the original Chinese (Persian) text, the second column is the Persian (Chinese) translation, and each row corresponds to one pair of entries. For the KB-augmented condition, we employed a structured instruction, directing selected models to: (i) produce accurate and natural Persian/Chinese that conforms to Persian (Chinese) language conventions, particularly in the use of euphemistic expressions; (ii) pay special attention to cultural-loaded content such as idioms, cultural allusions, character names, place names, and festival names; (iii) prioritize querying the internal knowledge base for referential translation, with emphasis on euphemisms and cultural equivalents; (iv) when knowledge base materials were insufficient, conduct online searches for relevant, verifiable English-Persian translation resources; and (v) if necessary, adopt an intermediate-language strategy, ensuring fluency, naturalness, cultural equivalence, and retention of euphemistic elements. The final instruction required the model to output only the Persian (Chinese) translation, without labels, explanations, or additional descriptive text.

5 COMPARATIVE ANALYSIS METHOD OF TAAROF TRANSLATIONS

The difficulties in translating *taarof* stem from culture-specific contexts and cultural norms unique to Persian communication. These contexts include host-guest interactions where no genuine offence has occurred - for instance, a host apologizing for “bad food” despite having prepared an elaborate meal. Such expressions are guided by *taarof*, emphasizing values such as *tavāzo* (humility), *mehmān-navāzi* (hospitality), and *shekasteh-nafsi* (self-deprecation). These norms create discrepancies between “what is said” (e.g., “the food is bad”) and “what is meant” (e.g., “I am being polite by downplaying my cooking”), posing significant challenges for translators attempting to convey the intended meaning to English-speaking audiences, where such apologies rarely occur without genuine offence [12].

In this study, we conducted two qualitative translation assessments involving Persian, Chinese, and English. The first assessment focused on a speech delivered by a native Persian-speaking teacher at a welcoming ceremony at a Chinese university, translated from Persian into Chinese. The second assessment examined a set of English subtitles translated first into Chinese and subsequently into Persian. For each assessment, five translation versions were produced using different methods: GPT-4o integrated with a self-constructed KB in Dify AI, DeepSeek-V3.1 with the same KB, and three independent LLMs, namely GPT, Claude, and WPS DeepSeek AI. The Chinese translations were evaluated by one of the authors, a Chinese lecturer specializing in Persian language instruction. The Persian translations were evaluated by the same author in collaboration with two native Persian speakers pursuing Chinese language studies at a

Chinese university. The self-customized KB comprises several peer-reviewed articles accessible online, focusing on politeness and euphemism research across Chinese, English, and Persian, written in either Chinese or English.

Table 1 Original Persian Speech Text and Corresponding English Literal Meanings

| Persian speech text | English literal meaning |
|------------------------------------|---|
| مهمانان گرامی | Esteemed guests |
| دانشجویان عزیز زبان فارسی | Dear students of Persian language |
| شب شاعرانه و صمیمی | A poetic and intimate night |
| دور هم جمع شدن | Gathering |
| جشن گرفتن | Celebrating |
| یکی از جشنهای باستانی | One of the ancient festivals |
| شب یلدا | Yalda Night |
| لحظه خاص | A special moment |
| به اشتراک گذاشتن لذت و برکت | Sharing joy and blessings |
| دو مورد شباهت | Two similarities |
| جشن دونگجی | <i>Dongzhi</i> Festival (Winter Solstice Festival in China) |
| سرزمین چین | The land of China |
| پایان طولانیترین شب | The end of the longest night |
| خوش آمد گویی به بازگشت نور | Welcoming the return of light |
| اسم متفاوت | Different names |
| نماد تمایل | Symbol of aspiration |
| نور | Light |
| گرمی | Warmth |
| دورهمی | Gathering |
| تصادف زمانی | Temporal coincidence |
| ارتباط روح | Spiritual connection |
| شاهد | Witness |
| آداب و رسوم آشنا | Familiar customs |
| کشتن گوسفند و خوردن جياوزه در شمال | Slaughtering sheep and eating <i>jiaozhi</i> in the north |
| خوردن توپ برنجی در جنوب | Eating rice balls in the south |
| آرد برنج چسبناک | Glutinous rice flour |
| توپ برنجی مخصوص دونگجی | Special <i>Dongzhi</i> rice balls |
| نماد اتحاد | Symbol of unity |
| آماده کردن غذاهای خوشمزه | Preparing delicious foods |
| قربانی کردن برای نیاکان | Making offerings to ancestors |
| مانند جشن سال نو | Like the New Year festival |
| انسجام ملی | National cohesion |
| تجربه شخصی | Personal experience |
| یازدهمین جشن یلدا | The eleventh Yalda celebration |
| آشنا شدن | Becoming acquainted |
| هرگز غافل نشدن | Never neglecting |
| همکلاسیهای زبان فارسی | Persian language classmates |
| همکاران زبان فارسی | Persian language colleagues |
| خانواده زبان فارسی | Persian language family |
| معنای خاص | Special meaning |
| زبان آموزان فارسی | Persian language learners |
| پل ارتباطی | Bridge of communication |
| پیوند بین تمدنها | Link between civilizations |
| ابزار زبانی | Linguistic tool |
| درک عمیق | Deep understanding |
| جوهره فرهنگ فارسی | The essence of Persian culture |
| به چین بازگرداندن | Bringing back to China |
| سفیر فرهنگی | Cultural ambassador |
| ترویج تبادلات | Promoting exchanges |
| مردم چین و ایران | The people of China and Iran |
| شناخت متقابل | Mutual understanding |
| گذشت زمان طولانی | The passage of a long time |
| دوستی عمیق | Deep friendship |
| دستاوردهای تبادل آموزشی | Achievements of educational exchange |
| افراد بیشتری | More people |
| علاقهمند به زبان فارسی شدن | Becoming interested in Persian language |
| یادگیری زبان باستانی و زیبا | Learning the ancient and beautiful language |
| ابراز علاقه کردن | Expressing interest |
| تبادل فرهنگی دوطرفه | Bilateral cultural exchange |
| دمیدن جان تازه | Breathing new life |
| پیشگامان توسعه فرهنگی چین و ایران | Pioneers of cultural development between China and Iran |
| تشکر صمیمانه | Heartfelt thanks |
| عشق به زبان فارسی | Love for the Persian language |
| تجربه جذابیت خاص | Experiencing a unique charm |

| | |
|-------------------------|--------------------------------------|
| پل دوستی | Bridge of friendship |
| ارتباط قلبی میان ملت‌ها | Heartfelt connection between nations |
| همدلی کامل | Mutual understanding and respect |
| پیشروی دست در دست | Moving forward hand in hand |
| فارسی زبانان | Persian speakers |
| در هر کجا که باشند | Wherever they may be |
| گرما و عشق | Warmth and love |
| چراغی فروزان | A glowing lamp |
| هرگز خاموش نشدن | Never to be extinguished |
| روشن کردن مسیر آینده | Illuminating the path ahead |
| به سرخی انار | As red as a pomegranate |
| به سرخی هندوانه | As red as a watermelon |
| خندهای چون پسته | A smile like a pistachio |
| عمر به بلندی شب یلدا | A life as long as Yalda night |
| شبی زیبا و فراموش نشدنی | A beautiful and unforgettable night |

Table 2 Scores of Five Chinese Translation Versions

| Chinese Translation (CT) Versions | Scores |
|---|--------|
| CT 1 (GPT-4o with a self-constructed KB in Dify) | 91.25 |
| CT 2 (Vanilla GPT-5-high chat model) | 86.25 |
| CT 3 (Vanilla claude-sonnet-4-5-20250929-thinking-32k chat model) | 83.75 |
| CT 4 (DeepSeek-V3.1 with the same KB) | 88 |
| CT 5 (WPS DeepSeek AI) | 88.75 |

Among the five automated Chinese translation (CT) versions, CT 1 stands out as a balanced and reliable performer, excelling in core accuracy while maintaining functional fluency. It correctly renders Chinese translations of key terms like Winter Solstice and glutinous rice flour with semantic equivalence. Overall, the translation avoids critical errors.

In contrast, other versions have distinct strengths and flaws: though being idiomatic, CT2 exhibits over-literal wording that misaligns with the rhetorical conventions of Chinese celebration speeches. For example, the Chinese phrase that literally means “never ignored” carries a negative connotation by highlighting the absence of neglect rather than the presence of proactive care, respect and affirmation; similarly, the phrase that literally means “sufficient empathy” feels quantitative, clashing with the colloquial language and joyous emotion expected in such positive contexts. These literal translations prioritize linguistic accuracy, resulting in a tone that is too formal, detached, or even slightly negative, contradicting the celebratory, uplifting atmosphere the speech aims to create.

CT 3 suffers from similar stylistic awkwardness, thereby fails to adapt to the contextual pragmatics of a festive setting. For instance, the phrase that literally means “psychological connection” is overly formal and abstract. Celebration speeches emphasize emotional closeness, so natural alternatives like “heartfelt/emotional bond/connection” would better evoke intimacy, whereas the former feels detached, like a philosophical term rather than a heartfelt tribute. Similarly, the phrase that literally means “Winter Solstice custom-made *tang-yuan*” includes the redundant qualifier: *tang-yuan* is inherently associated with Winter Solstice in Chinese culture, so adding “custom-made” implies the repetitive language use, diverging the focus on its symbolic meaning of unity and warmth; a more natural phrasing would be as explicit as “Winter Solstice *tang-yuan*”, which leans into the dish’s cultural significance without unnecessary, jarring, and marketing-like language.

As to CT 4 and 5, one performs slightly better in contextual awareness (linking “Yalda Night” to “Winter Solstice” via parenthetical explanation) but can be apparently verbose. The other is undermined by a critical error - mistranslating “Ancestor worship” as “sacrifice our lives for relatives”, which distorts the cultural meaning.

To conclude, above five translations as shown in **Table 1** and **Table 2** demonstrate uneven performance in converting Persian *taarof* into Chinese euphemisms and polite expressions, with key limitations across accuracy, cultural adaptability, and naturalness. According to human evaluation, CT 1 (the use of an adequate LLM plus a KB) achieves the optimal balance between naturalness and semantic accuracy, emerging as the strongest in terms of general performance. However, the results still highlight the need for human translators to capture linguistic and cultural nuances that automated tools (even with a specially made KB) currently struggle to replicate.

Table 3 Original English Subtitles that are Converted to Chinese and then Persian Versions

| Sources of Video Content (game/TV series) | No. | English Subtitles |
|--|-----|--|
| <i>Divinity: Original Sin 2</i> | 1 | And then, when we made love, all of me... awoke. |
| | 2 | My daddy once said, “don’t go to the toilet on your own doorstep”. |
| | 3 | “It’s a cat-has-trophy.” Cat-has-trophy. Catastrophe. It’s a catastrophe. |
| | 4 | Hail, hail, hail... |
| <i>Grey's Anatomy</i> | 5 | I suppose that’s a death sentence for Fort Joy. What a terrible shame. |
| | 6 | Look at her belly. She is almost as wide as she is tall. |
| <i>Friends</i> | 7 | Listen to me! When my time comes, I wanna be buried at sea. |
| | 8 | It's like when you're a kid, and your parents put your dog to sleep, and they tell you it went off to live on some farm. |

| | | |
|-----------------------------|----|--|
| <i>Boyhood</i> | 9 | Smantha, why don't you say goodbye to that little horseshit attitude, okay? Cause we're not taking that in the car. |
| <i>Friends</i> | 10 | sorry I acted like a bank. |
| | 11 | I was laid to rest on Monday. |
| | 12 | After the funeral, all the residents of Wisteria Lane came to pay their respects. And as people do in this situation, they brought food. |
| | 13 | I don't need help. You hear those cheers? Those are for me. Last game. I threw 246 yards. With stats like that, I can get into the college of my choice. |
| <i>Desperate Housewives</i> | | Mr. Shaw: "No. Edie Britt will disappear, and you'll go back to your life." |
| | 14 | Carlos: "And what if she dies? People slip into comas all the time and never come out." |
| | 15 | Bree: I think you should know there's been an incident with your father. |
| | 16 | Carlos: Now even if she's...heavy, doesn't she have the right to feel good about herself? |
| | | Carlos: I know she's gotten chubby. She'll grow out of it. |

Regarding the five Persian translations (PTs) presented in **Table 3**, three versions (PT 2, 4, and 5) contain untranslated Chinese words within the Persian output, which should be considered a critical error of translation quality. The five translations respectively employ: PT 1 - GPT-4o with a self-constructed KB in Dify AI; PT 2 - GPT-5.1-high chat model; PT 3 - Claude-opus-4-5-20251101-thinking-32k chat model; PT 4 - DeepSeek-V3.1 with the same KB; and PT 5 - WPS DeepSeek AI. When comparing PT 1 and PT 3, evaluators expressed slightly differing opinions. One author - a Chinese university lecturer specializing in Persian language instruction - and two native Persian students studying Chinese at a Chinese university participated in the evaluation.

The author scored PT 1 at 90 and PT 3 at 95 out of 100, while rating the remaining translations below 80. She considers PT 1 to demonstrate overall fidelity with minor stylistic loss, whereas PT 3 exhibits semantic completeness and superior naturalness, though with excessive free translation. For example, for the sentence "don't go to the toilet on your own doorstep", PT 1 renders it as "do not cause trouble where you live" while PT 3 renders it as "do not vandalize your own house". Both translations interpret the euphemistic phrase "go to the toilet" which implies "pee" or "urinate" though PT 1 appears to obscure the original meaning.

From the perspective of the two native Persian speakers, they consider the first translated sentence in PT 3 to be impolite: بعد از آن، وقتی پوستمان به هم رسید، همه چیز در وجودم... بیدار شد. The female evaluator found it offensive, whereas her husband did not express strong emotions about it. Regarding this sentence, she gave high scores only to PT 1 and PT 4, both of which employ the KB, while considering the outputs from vanilla LLMs to be rude and offensive. Interestingly, for sentence 8, both native Persian speakers fully agreed with PT 1 and identified errors in PT 3. However, the Chinese evaluator maintained that both translations are overly polite and expressed a preference for an explicit term of "اتنازی" rather than a euphemistic and indirect explanatory rendering. In this regard, the other two authors aligned with the original English text, arguing that this euphemistic style should be preserved in the target text. Overall, PT 1 has an average error rate of 28.13%, whereas PT 3 has an error rate of 31.25%. We conclude that, considering the combination of semantic accuracy and politeness transfer, integrating a KB into the automated translation process is advisable for contexts requiring strict translation quality. Additionally, results from both translation directions (Persian - Chinese and English - Chinese- Persian) indicate that the KB-augmented GPT model outperforms the KB-augmented DeepSeek model. This suggests that the choice of underlying LLM for KB integration is itself a crucial factor in determining overall translation quality.

6 LIMITATIONS

Unlike high-resource language pairs such as translating between English and Chinese, which benefit from well-established evaluation frameworks employing standardized criteria including accuracy, fidelity, fluency, naturalness, and cultural appropriateness, LRLs such as Persian lack unified assessment standards. Consequently, the three evaluators in this study - one Chinese lecturer and two native Persian speakers - scored translations based primarily on individual linguistic intuitions rather than a rigorously defined rubric. Although semantic accuracy, naturalness, and politeness transfer were considered, the weighting of these criteria varied among evaluators, inevitably reflecting personal preferences and cultural biases regarding politeness norms and euphemism interpretation. This lack of standardized protocols represents both a limitation of this study and a broader gap in the field warranting future attention.

Additional constraints concern the limited scope of the self-constructed KB, which may not comprehensively capture politeness and euphemism conventions across all three languages. The KB currently consists only of a set of formally published academic articles on these topics written in English or Chinese, which restricts the representation of culture-specific Persian practices and may limit the extent to which the models can effectively exploit the external knowledge. Moreover, the translation samples were confined to two text types, namely ceremonial speech and audiovisual subtitles, which may reduce the generalizability of the findings to other genres or communicative situations. In addition, we did not systematically control for or analyze differences in prompting strategies between LLMs used with and without the KB, which may also have influenced performance. Finally, the rapid evolution of LLMs suggests that the results may not remain fully applicable to future versions of these systems.

7 CONCLUSION

This study summarized features of euphemistic expressions and investigated the challenges of accurately and appropriately translating nuances of FTAs and politeness expressions between Persian, Chinese, and English, with particular attention to the Persian cultural schema of *taarof*. Through qualitative assessment of translations produced by various automated tools, we examined how effectively these systems capture linguistic and cultural nuances inherent in politeness markers and euphemistic expressions.

Our findings reveal that translations generated by LLMs integrated with a custom-built KB consistently outperformed those produced by standalone LLMs in terms of semantic accuracy and appropriate politeness transfer, particularly when an appropriate LLM is combined with the KB. The inclusion of domain-specific knowledge pertaining to politeness conventions and euphemism usage across the three languages demonstrably enhanced translation quality, reducing error rates and improving cultural appropriateness. However, even the best-performing automated translations exhibited limitations in fully capturing the nuanced interplay between what is said and what is meant in *taarof* contexts, where ritualized politeness often diverges significantly from literal interpretation.

The divergent evaluations between native Persian speakers and the Chinese evaluator further underscore the complexity of cross-cultural translation assessment. What one cultural group perceives as appropriately polite may be interpreted as overly indirect or even offensive by another. Besides, KB achieves varied performance of translation quality on the base of different LLMs. Our finding reinforces the argument that politeness is not a universal construct but rather a culturally contingent phenomenon that poses significant challenges for automated translation systems. While advancements in LLMs have substantially improved automated translation capabilities, human involvement is still a necessity for the translation of culturally embedded speech acts such as requests, apologies, and euphemisms. We recommend that translation practitioners and developers consider integrating specialized KBs when translating between languages with distinct politeness systems prior to post-editing.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

FUNDING

This work was supported by the Guangdong Philosophy and Social Science Research Programme (GD25CWY20) and the Guangdong Postgraduate Education Reform Programme (2025JGXM_064).

REFERENCES

- [1] Dong Xiaojun. A Discussion on the Communicative Functions of Euphemisms in English Subtitle Translation: A Case Study of the British Drama Downton Abbey. English Square, 2016(10): 5-7. DOI:10.16723/j.cnki.yycg.2016.10.002.
- [2] Naumenko S, Kryvonis M. Linguistic means of forming euphemisms in the modern Chinese language. Cognition, Communication, Discourse, 2023(27): 44-55.
- [3] Kamehkhosh N, Larina T V. Cultural values and politeness strategies in British and Persian family discourse. Proceedings of INTCESS, 2020: 7th.
- [4] Tayebi T. Heterogeneous distribution of cultural conceptualizations and (im) politeness evaluations. International Journal of Language and Culture, 2020, 7(1): 84-103.
- [5] Sharifian F, Tayebi T. Perception of (im) politeness and the underlying cultural conceptualisations: A study of Persian. Pragmatics and Society, 2017, 8(2): 231-253.
- [6] Kádár D Z, House J, Liu H. Beyond the deferential view of the Chinese V pronoun *nin*. Pragmatics, 2025, 35(2): 155-184.
- [7] Yaghoubzadeh S. A study of requests made by negative politeness in English movies and their Persian subtitles. Theory and Practice in Language Studies, 2021, 11(2): 216-223.
- [8] Zohreh Ameri Safiabadi. A Comparative Study of Politeness Formulae in Chinese and Persian. Zhejiang University, 2024. DOI:10.27461/d.cnki.gzjdx.2024.001957.
- [9] Ji Kaiyun, Yang Chen Ying. A Cultural Interpretation of Iranian Taarof from the Perspective of Cultural Anthropology. Middle East Studies, 2021(01): 147-164+265-266.
- [10] Izadi A. Over-politeness in Persian professional interactions. Journal of Pragmatics, 2016, 102: 13-23.
- [11] Wang X, Yang Q, Qiu Y, et al. Knowledgpt: Enhancing large language models with retrieval and storage access on knowledge bases. arXiv preprint arXiv:2308.11761, 2023.
- [12] Yaqubi M. On subtitling of ta'arof apologies. Journal of Language and Translation, 2018, 8(1): 31-42.