The Construction of the "Sinolingo" Platform and a Study on Its Effectiveness in Autonomous Translation Learning

Mingkai Li[1,a], Qiaoxi Huang[1,b], Jinwen Hu[1,c], Renjia Chen[1,d], Menglian Liu[2,e], Xingyu Pu[1,f*]

[1]School of Interpreting and Translation Studies, Guangdong University of Foreign Studies, Guangzhou, China
[2]Laboratory of Language and Artificial Intelligence, Guangdong University of Foreign Studies, Guangzhou, China

[a]liamlmk06@gmail.com, [b]734198330@qq.com, [c]2782122078@qq.com, [d]1342924479@qq.com, [e]lmenglian@gdufs.edu.cn, [f*]2195686045@qq.com

Abstract. This study investigates the efficacy of "Sinolingo", a customized translation platform designed for traditional Chinese culture and powered by DeepSeek, in supporting autonomous translation learning. The paper elaborates on the platform's design philosophy and core architecture, introducing the "TRICE" framework. Through empirical research, it demonstrates that the platform significantly enhances autonomous translation learning, reflected in the effectiveness of both the overall platform and its individual functional components. The results indicate that the platform effectively facilitates translation learning, enhances the quality of students' output, and provides a valuable reference for the design and implementation of intelligent computer-assisted translation tools in pedagogical contexts.

Keywords: Autonomous Translation Learning, Large Language Models (LLMs), Computer-Assisted Translation Pedagogy, Translation of Chinese Culture, Prompt Engineering

1 Introduction

1.1 Research Background and Motivation

As the "Chinese Culture Going Global" initiative advances, the demand for highly qualified translators has become increasingly urgent. Translation is not merely a linguistic endeavor; it is also an essential medium for cultural dissemination and the construction of national identity [1]. Nevertheless, translation education continues to encounter enduring challenges: fragmented learning resources limit students' access to comprehensive input, the lack of timely and effective feedback impedes progress, and the intricacies of cultural nuance exacerbate comprehension and expression [2][3].

A. Y. M. A. Islam et al. (eds.), *Proceedings of the 2025 International Conference on Educational Technology and Management Information Systems (ETMIS 2025)*, Advances in Computer Science Research 129,
https://doi.org/10.2991/978-94-6239-630-2_54

The emergence of large language models (LLMs) has opened new avenues for enhancing translation pedagogy. Xu, Sun, Su, and Liu [3] contend that these models demonstrate superior capabilities in natural language comprehension and generation, allowing them to serve not only as translation tools but also as platforms that foster learner autonomy through prompt engineering and tailored feedback. Furthermore, systematic reviews like AlTwjiri and Alghizzi show that the integration of AI in education can boost students' engagement and motivation while facilitating learning[4]. In the context of translation pedagogy, Xu, Su, and Liu present evidence that AI-integrated platforms combining translation, post-editing, and feedback can cultivate comprehensive learning environments that foster reflective practice and skill enhancement[5].

1.2 Research Objectives

The primary objective of this study is to develop and evaluate "Sinolingo," a customized translation platform for traditional Chinese culture leveraging LLMs, and to assess its effectiveness as a tool for autonomous translation learning. The platform was created to help with ongoing problems in translation education, such as disjointed resources, insufficient feedback, and difficulties in interpreting and conveying cultural meaning. "Sinolingo" wants to give learners real, context-sensitive translation help that goes beyond what regular tools can do by combining LLM technology with domain-specific corpora and prompt engineering.

The study also aims to foster learner autonomy by encouraging engagement with key platform features, including prompt templates, corpus retrieval, and translation memory. These features are intended to promote critical thinking, self-reflection, and improved translation output. Furthermore, the project seeks to enhance the domain of AI-assisted language learning by providing empirical evidence regarding the educational efficacy of customized platforms and by introducing the TRICE prompt engineering framework (Task, Requirement, Information, Corpus, Evaluation) as a replicable model for analogous educational settings.

1.3 Research Questions

Based on these goals, the study addresses two main research questions.

The first examines the platform's overall effectiveness compared to traditional translation tools: Does "Sinolingo" significantly enhance the quality of students' translations in terms of accuracy, fluency, and cultural fidelity? This question aims to determine whether the integration of LLMs with structured pedagogical guidance can lead to measurable improvements in translation outcomes.

The second investigates the role of specific platform features in supporting autonomous learning: Is there a positive correlation between learners' use of prompt templates and corpus retrieval and their translation performance? This inquiry seeks to identify not only the overall utility of the platform but also the specific components that contribute most to learner development.

Together, these questions provide a framework for evaluating the platform's pedagogical impact and the mechanisms by which it fosters translation competence.

2 Literature Review

2.1 CAT Platforms and Translation Teaching

The integration of Computer-Assisted Translation (CAT) platforms into translation pedagogy is a key research direction in natural language processing. AI-powered systems, despite their strengths in semantic understanding and contextual processing, still fall short in Chinese-to-English translation of traditional Chinese cultural texts[6]. Compared with human translators, current AI systems have gaps in accuracy, fluency, and cultural adaptation[7], and even commercial CAT platforms with AI-generated content (AIGC) tools face issues of inconsistent and unstable translation quality[8].

In general, natural language processing tasks, AI models perform well, but they show obvious weaknesses when dealing with texts containing deep cultural meanings. For tasks requiring cultural knowledge, they often need additional technical support to work effectively[9], especially in translating Chinese philosophical texts, where they struggle to fully convey cultural concepts and subtleties.

To address these problems, researchers are exploring solutions such as multimodal fusion and the construction of specialized linguistic databases, which have shown initial potential in improving the translation of cultural connotations in traditional texts[10]. However, existing efforts still have room for improvement, as some attempts to optimize cultural text translation still suffer from issues like unnatural phrasing and incomplete preservation of cultural imagery. In summary, the application of CAT platforms in translation education and practice is still in the development stage. For traditional cultural translation specifically, there is an urgent need for progress in key areas such as the construction of specialized corpora, the integration of cultural context, and the advancement of multimodal methodologies.

2.2 The Use of AI Platforms in Self-Directed Translation Learning

AI platforms applied to self-directed translation learning have attracted increasing research attention, particularly due to their potential to adapt to specific subject areas, a feature that holds significant value for targeted translation learning.

Internationally, initial progress has been made in optimizing AI translation systems, such as developing functions that allow users to customize translation styles. However, these optimization processes face challenges like high reliance on computational resources, leading to high costs and difficulty in large-scale promotion. Additionally, while some major technology companies have launched projects focusing on culturally sensitive translation, they often fail to accurately capture and express the profound philosophical concepts unique to Chinese thought[11]. However, research by Peng et al. also shows that LLMs are heavily dependent on the cultural content of their training data, which limits their effectiveness in low-resource cultural settings[12].

In China, research in this field has focused more on developing terminology adaptation systems for modern professional fields (e.g., law and medicine), which enhance translation quality by integrating expert knowledge from specific domains. However, most of these platforms are not well-suited for traditional cultural contexts, as they do not fully consider the unique linguistic styles and cultural imagery in traditional texts. Although some platforms have made progress in corpus development and technical innovation, their performance in traditional cultural translation still lacks consistency in aspects like supporting autonomous learning and generating adaptive outputs.

3 Design of the Platform

3.1 Overall Platform Architecture

This study relies on the large language model DeepSeek, combined with a self-built Chinese–English bilingual corpus of Chinese traditional culture and TRICE prompt engineering framework, to construct a customized translation platform for cultural texts, namely “Sinolingo”. The platform is designed not only to provide practical translation functions but also to serve as an experimental system for research validation, thereby exploring the potential of integrating prompt engineering with domain-specific corpora in cultural translation scenarios.

As illustrated in Figure 1, the platform is structured around three main layers: user interaction, frontend display, and backend management. These layers cover the workflow from source text input and prompt customization to translation generation, evaluation, and refinement. In addition, customized workflows, AI self-checking, and translation tools provide horizontal support across different stages of the process. This design further highlights the interaction between pedagogical modules such as standardized corpora, terminology databases (hereafter referred to as termbase/termbases), and translation memories, together with backend processes, thus providing a clear framework for linking translation pedagogy with technological support.

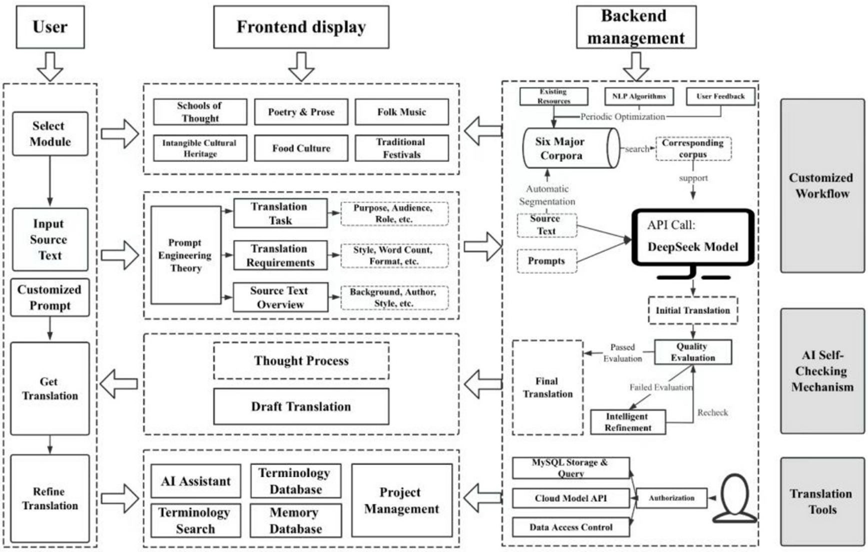

Fig. 1. Overall architecture of the platform.

3.2 TRICE Framework and Customized Translation Mechanism

To ensure both the practical implementation of customized translation and the verifiability of research outcomes, the platform adopts the TRICE prompt engineering framework, which integrates five interrelated components. Task defines the nature of translation, specifying domains such as academic, literary, or popular science, thereby shaping register and rhetorical choices. Requirement refines stylistic and communicative parameters by configuring tone, audience orientation, and stylistic goals (e.g., maintaining a classical style or enhancing accessibility), thereby ensuring alignment with the intended readership. Information supports the process through prompt templates and termbases, which provide structured guidance and domain-specific terminology; for instance, rhyme prompts are applied in poetry, while specialized termbases support technical fields such as medicine, thereby improving contextual adequacy. Corpus ensures adaptation and consistency in terminology and cultural features through a self-constructed bilingual corpus, thereby enhancing the reliability of subsequent evaluation. Evaluation combines automated metrics (BLEU, chrF++, BERTScore) with expert human judgment to balance quantitative assessment and qualitative insights.

Preliminary feedback indicated that the TRICE workflow helped preserve cultural imagery, supported more accurate terminology use, and promoted greater stylistic consistency. Compared with baseline conditions, translations produced under TRICE were observed to better reflect the features of traditional cultural texts and to show improvements in readability and professionalism, as shown in Figure 2.

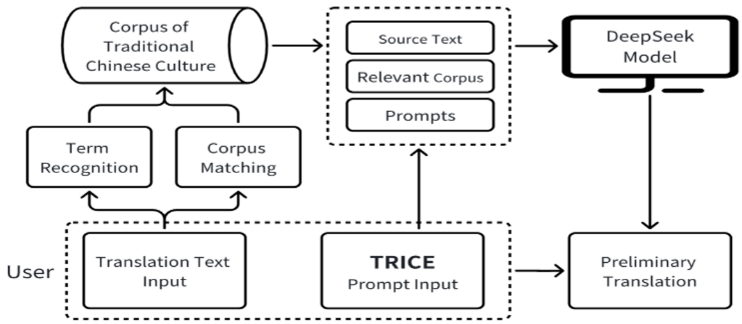

Fig. 2. Application flow of the TRICE framework in the platform.

3.3 Functional Modules and Implementation Outcomes

Translation Main Page and AI Assistant. The translation main page is the platform’s core workspace, supporting Quick Translation for baseline comparisons and Customized Translation with templates and termbases for research-oriented tasks. Its layout prioritizes simplicity and experimental controllability, with navigation functions enabling rapid switching between resources. Version management ensures traceability by preserving outputs under different conditions of template use and terminology operations, thus providing comparable data samples. Testing indicated that Quick Translation maximizes efficiency, while Customized Translation better conveys cultural imagery and maintains consistency. Feedback confirmed that the main page supports both daily translation tasks and controlled experimental designs.

In addition, a sidebar AI Assistant is embedded into the main page as an auxiliary module. Rather than functioning as a standalone page, it provides on-demand refinement tools such as style harmonization, terminological unification, and background knowledge supplementation. This design enables users to optimize translations without leaving the workflow, particularly benefiting poetic or literary texts where subtle adjustments are crucial. Although not a direct variable in hypothesis testing, the assistant effectively reduces post-editing workload and enhances the platform's practical utility, ensuring that outputs are not only research-ready but also polished for cross-cultural communication. As shown in Figure 3.

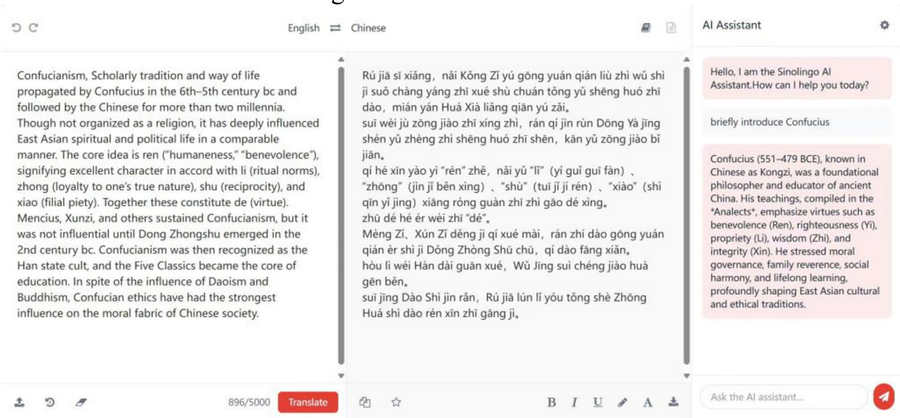

Fig. 3. Interface of the Translation Main Page.

Template and Termbase Management. Figure 4 shows the interface of the template management module. In alignment with the TRICE framework, the platform implements template and termbase management as its two core modules for prompt engineering and terminology control. The template module allows researchers to create, edit, delete, and categorize templates for systematic prompt engineering. Templates are organized into six thematic categories (e.g., "Schools of Thought," "Traditional Festivals") and may include variable slots such as {style} or {audience}, allowing researchers to flexibly adjust translation parameters. For instance, within the same template framework, one may select an academic style for scholars or a poetic style for general readers, enabling parallel outputs from a single source text under varied experimental conditions. This design externalizes task analysis, supports structured pre-translation planning, and introduces controllable experimental conditions. In doing so, it enhances the comparability and reproducibility of translation experiments, making it possible to observe how different degrees of template use affect translation processes and outcomes.

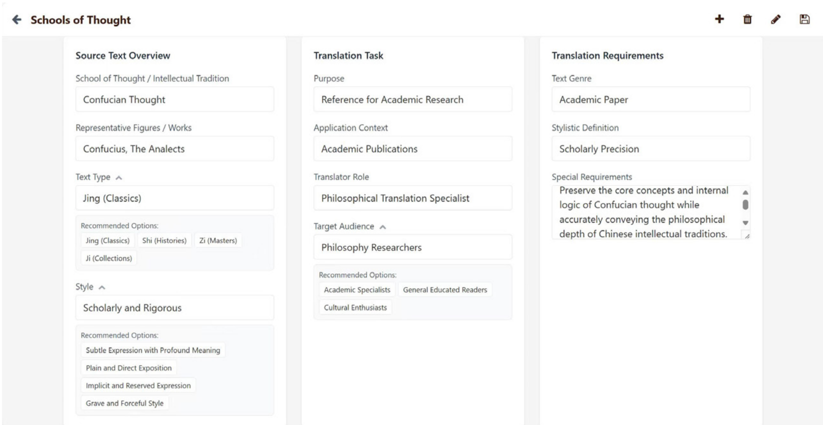

Fig. 4. Interface of Template Management.

The termbase module supports terminology control through keyword lookups, on-screen highlighting, and real-time match display with one-click unified replacement. During translation, researchers can instantly view suggested terms and standardize their application, ensuring consistency across outputs and leaving traceable records of terminology use for experimental analysis. This improves reliability in specialized domains and strengthens result comparability. The translation memory, though not central to experimentation, offers sentence-level reuse that reduces repetitive work and stabilizes performance in long-text processing.

4 Methodology

4.1 Theories and Assumptions

Translation Platform Use and Translation Quality Improvement. Research in the Learning Sciences has shown that the design of learning environments significantly influences learners' cognitive processes and outcomes [13]. The translation learning platform, built on the TRICE framework, provides task orientation, corpus support, and process feedback, offering structured cognitive scaffolding to optimize information processing and problem-solving, thereby improving learning efficiency and outcomes[14].

At the same time, Self-Regulated Learning (SRL) theory emphasizes that learners engage in goal setting, strategy selection, progress monitoring, and self-reflection during the learning process [15]. Research has shown that the value of learning tools lies not only in providing resources but also in helping learners clarify task requirements and enhance metacognitive regulation [16]. The prompt template function of Sinolingo platform externalizes task analysis and planning, strengthening learners' goal-setting and strategy development while reducing cognitive load associated with unclear objectives [17][18]. In addition, corpus retrievals (operationalized as termbase retrievals in the Sinolingo platform) and terminology matches provide verifiable and immediate linguistic resources, which reduce the cognitive burden of lexical searching, comparison, and revision, thus improving the accuracy and fluency of translations. Therefore, under the support of Learning Sciences and SRL theory, translation platforms are expected to significantly improve students' translation quality.

Hence, the following hypothesis is proposed:

H1: The use of the translation learning platform significantly improves students' translation quality.

The Role of Key Platform Functions. SRL theory further indicates that learning outcomes depend not only on whether learners use a tool but also on the depth and breadth of their engagement [19][20]. The completeness of prompt template filling reflects the learner's level of engagement during task preparation: a higher degree of completeness suggests a more thorough analysis of task requirements and more effective strategic regulation, which in turn improves translation performance [18]. Thus, template completeness can be regarded as an external indicator of metacognitive regulation.

In addition, frequent termbase retrievals and terminology matches reflect learners' resource management ability during translation. Although research on termbase use in translation pedagogy is relatively limited and often focuses on terminology management and consistency in professional contexts [21][22], the pedagogical principles are analogous to those found in corpus-assisted learning. Both corpora and termbases provide structured external resources that support learners' metacognitive regulation and resource management strategies. Prior research in corpus-based learning demonstrates that proactive resource use is positively associated with learning outcomes, particularly in complex tasks such as language learning and translation [23]. These insights provide transferable evidence to justify the role of termbase functionalities in improving translation quality.

Hence, the following hypotheses are proposed:

H2: The completeness of prompt template filling has a significant positive effect on translation quality.

H3: The number of termbase retrievals and terminology matches has a significant positive effect on translation quality.

4.2 Participants

The participants of this study were thirty undergraduate students majoring in translation, randomly selected from the same institutional setting. All participants had received prior training in translation theory and practice, ensuring a comparable baseline of competence. Random sampling was employed to minimize potential bias in quality differences and to strengthen the validity of subsequent statistical analysis. Participation was voluntary, and all students were informed of the study's purpose and procedures before providing consent.

4.3 Research Design

The study adopted a quasi-experimental design to evaluate the effectiveness of the Sinolingo platform and to test the hypotheses formulated in Section 4. Three translation tasks were designed for the thirty participants. In Task 1, each student translated a source text using conventional methods, such as dictionaries or general-purpose LLMs. The outputs were labeled S1–S30, where S denotes "self" and the number indicates the student (e.g., S1 refers to the first student's translation using traditional tools). In Task 2, the same group translated another text using the Sinolingo platform under the TRICE framework. The outputs were labeled P1–P30, where P denotes "platform" (e.g., P1 is the first student's translation produced with Sinolingo). Finally, in Task 3, students used the platform autonomously for one month before translating a new text again with the support of Sinolingo. The outputs of this final stage were labeled T1–T30, where T denotes "training."

To ensure experimental validity and comparability across tasks, all three translation tasks were designed under a controlled-variable framework. Specifically:

Topic control: The source texts for the three tasks were drawn from comparable domains and discourse types, including both general informational texts and culturally loaded materials such as classical Chinese texts, in order to control for variation in interpretive complexity. This design ensured that differences in translation performance were not attributable to topic familiarity alone but also reflected the ability to handle culturally embedded meaning. Length control: Each text contained approximately the same number of words (around 300 ± 10 words), ensuring comparable workload across tasks. Terminology density: The number of domain-specific or culturally salient terms embedded in each text was kept consistent, allowing a fair assessment of terminology-handling ability. Lexical richness: Vocabulary diversity (e.g., measured by type-token ratio, TTR) was balanced across texts, avoiding bias due to lexical variation. Readability: Text readability was controlled using standard indices (e.g., Flesch–Kincaid grade level) to maintain similar levels of syntactic and cognitive complexity.

The overall research design is summarized in Table 1.

To test H1, translations from Task 1 and Task 2 (30 pairs) were evaluated against reference translations using BLEU, chrF++, and BERTScore. Mean scores were compared and a paired-samples t-test assessed statistical significance.

Table 1. Research Design Overview.

Task	Description	Output Label	Statistical Test
Task1	Translate source text using conventional methods(dictionary + general LLMs)	S1-S30	Paired *t*-test (vs. Task 2)
Task2	Translate source text using Sinolingo platform under TRICE framework	P1-P30	Paired *t*-test (vs. Task 1)
Task3	Translate source text after 1 month of autonomous use of Sinolingo	T1-T30	Pearson correlation with usage logs

To test H2 and H3, platform usage data were recorded automatically during the one-month autonomous learning phase. The logs tracked the frequency of three key behaviors: completeness of prompt template filling, termbase retrievals, and terminology matches. Pearson's correlation coefficient (*r*) was then calculated to examine whether the frequency of these behaviors was positively associated with translation quality in

Task 3 (T1–T30). This analysis provided insights into the extent to which specific functionalities of the platform contributed to learning outcomes.

4.4 Data Collection

Translation Quality Data. Translation quality data came from three tasks (T1–T30, S1–S30, P1–P30). For Task 1 and Task 2, thirty pairs of translations (S1–S30 vs. P1–P30) were compared with a reference translation using BLEU, chrF++, and BERTScore to measure lexical overlap, character-level similarity, and semantic adequacy. Scores were averaged and tabulated to enable direct comparison. To further test H1, all scores were standardized, and a paired-samples t-test determined whether differences across the three metrics were statistically significant . The results of this analysis are reported in the Results section (see Table 2).

In addition, outputs from Task 2 and Task 3 (P1–P30 vs. T1–T30) were evaluated with the same three metrics. Averaged and tabulated results examined whether one month of continued Sinolingo use produced sustained improvements in translation quality, providing evidence of its long-term pedagogical benefits. The corresponding results are reported in the Results section, as shown in Table 5.

Platform Usage Data. In addition to translation performance, the system automatically recorded platform usage data during the one-month autonomous learning phase. Specifically, the logs tracked the frequency of three types of learner behaviors: the completeness of prompt template filling, the number of termbase retrievals queries, and the number of terminology matches. These variables were used as indicators of how actively students engaged with the core functionalities of Sinolingo.

For template completeness, the system calculated a percentage score based on the proportion of fields filled out within the TRICE framework. The formula was:

$$\text{Completeness}(\%) = \frac{\text{Number of fields filled}}{\text{Total fields}} \times 100\% \tag{1}$$

For example, if a student filled out 4 of 5 fields, the completeness score would be 80%.

To test H2 and H3, Pearson's correlation coefficient (r) was employed to examine whether the completeness of prompt template filling and the frequency of the other two behaviors were significantly associated with translation quality in Task 3(T1–T30). This analysis provided insights into the extent to which frequent use of key platform functions contributed to improved performance. The results offered evidence that sustained and systematic use of Sinolingo was positively correlated with translation ability, thereby demonstrating the platform's long-term pedagogical advantages. The results of this analysis are presented in the Results section, as shown in Table 5.

4.5 Power Analysis

To address concerns regarding the relatively small sample size (n = 30), a post-hoc power analysis was conducted to evaluate whether the statistical tests employed in this study had sufficient power to detect meaningful effects. The analysis focused on the primary hypothesis (H1), which was tested using paired-samples t-tests to compare translation quality scores between Task 1 (conventional tools) and Task 2 (Sinolingo platform).

Power analysis was performed using G*Power 3.1, with the test family set to "t tests" and the statistical test specified as "Means: Difference between two dependent means (matched pairs)." The significance level was fixed at $\alpha = 0.05$, in line with conventional standards in applied linguistics and educational technology research. The observed effect sizes (Cohen's d), derived from the empirical results reported in the Results section, ranged from medium to large (d = 0.62–0.78).

Based on these parameters, the achieved statistical power exceeded the commonly accepted threshold of 0.80. This indicates that the sample size was sufficient to reliably detect the observed effects and that the likelihood of Type II errors was adequately controlled. Therefore, despite the modest sample size, the statistical conclusions drawn from the paired-samples t-tests are supported by adequate power.

5 Result

This section presents the empirical findings derived from the three translation tasks and platform usage logs. The analyses are organized according to the research hypotheses (H1–H3). First, we compare translation quality between conventional tools and the Sinolingo platform, followed by an examination of long-term effects. Finally, we analyze the relationship between platform usage behaviors (template completeness, termbase retrievals, and terminology matches) and translation performance.

5.1 Effect of the Sinolingo Platform on Translation Quality

Hypothesis 1 (H1) was supported: the Sinolingo platform significantly improved students' translation quality compared with conventional tools. To test this, translation outputs from Task 1 (S1–S30) and Task 2 (P1–P30) were evaluated using BLEU, chrF++, and BERTScore. Table 2 summarizes the mean values of these three metrics for the two tasks.

As shown in Table 2, Task 2 scores were consistently higher across all three measures. This suggests that the platform enhanced both surface-level lexical accuracy and deeper semantic fidelity. To further examine whether these improvements were statistically reliable, paired-samples *t*-tests were conducted. The detailed results are reported in Table 3.

Table 3 shows significant improvements ($p < 0.05$) across all three metrics, indicating that students using "Sinolingo" achieved higher translation quality than with conventional tools. Effect size analysis (Cohen's d = 0.62–0.78) further confirms both statistical and practical significance, validating H1. A visual comparison of translation quality between Task 1 and Task 2 is provided in Figure 5.

In addition, a post-hoc power analysis was conducted for the paired-samples t-tests. Based on the observed effect sizes (Cohen's d = 0.62–0.78) and a significance level of $\alpha = 0.05$, the achieved statistical power exceeded the commonly accepted threshold of 0.80. This indicates that the sample size was sufficient to reliably detect the observed effects.

Table 2. Comparison of Translation Quality in Task 1 and Task 2.

Task	BLEU	ChrF++	BERTScore
Task1	21.4	48.7	0.812
Task2	28.9	56.2	0.864

Table 3. Paired-samples t-test Results for Task 1 vs. Task 2.

Metric	Mean (Task 1)	Mean (Task 2)	Mean Differ-ence	t value	p value
BLEU	21.4	28.9	+7.5	4.27	0.0002
chrF++	48.7	56.2	+7.5	3.95	0.0004
BERTScore	0.812	0.864	+0.052	3.41	0.0015

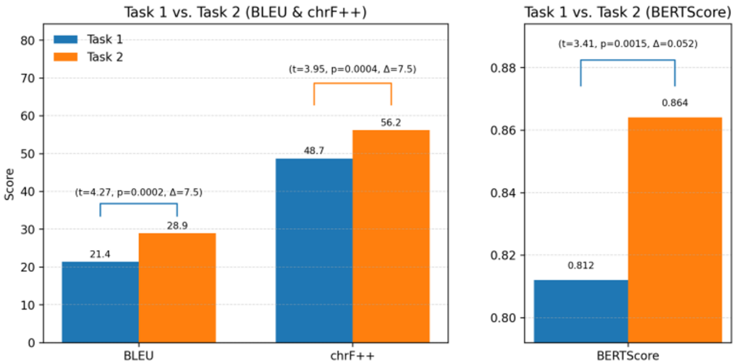

Fig. 5. Comparison of Translation Quality Between Task 1 and Task 2

5.2 Long-term Effects of Continued Platform Use

Beyond H1, an additional analysis compared Task 2 (immediate platform use) and Task 3 (after one month of autonomous platform use) to examine long-term effects. The same three evaluation metrics—BLEU, chrF++, and BERTScore—were applied. Table 4 presents the mean scores for the two tasks.

Table 4. Comparison of Translation Quality in Task 2 and Task 3.

Task	BLEU	ChrF++	BERTScore
Task2	28.9	56.2	0.864
Task3	31.7	59.4	0.881

As shown in Table 4, Task 3 outperformed Task 2 across all metrics, reflecting consistent improvements in translation quality. Although the gains were smaller than the sharp rise from Task 1 to Task 2, the upward trend underscores the platform’s sustained pedagogical value. These results suggest that continued autonomous use of “Sinolingo” consolidates students’ competence over time, leading to incremental but meaningful long-term gains.

5.3 Impact of Template Completeness on Translation Quality

Hypothesis 2 (H2) was supported: template completeness was positively associated with translation quality. During the autonomous learning phase (Task 3), each student’s template completeness score was correlated with their translation performance. As shown in Table 5, template completeness exhibited significant positive correlations with BLEU ($r = 0.52, p < 0.05$), chrF++ ($r = 0.49, p < 0.05$), and BERTScore ($r = 0.45, p < 0.05$).

These results indicate that students who filled out more fields—particularly those in the “translation requirements” category—produced translations with higher lexical overlap, better character-level accuracy, and stronger semantic adequacy. The findings suggest that structured use of the TRICE template provided cognitive scaffolding that directly contributed to improved translation outcomes.

Table 5. Correlation Between Platform Usage Frequency and Task 3 Translation Performance.

Platform Function	Measurement (per student)	Correlation with BLEU (r)	Correlation with chrF++ (r)	Correlation with BERTScore (r)	Significance (p)
Prompt template completeness	percentage of fields completed	0.52	0.49	0.45	<0.05
Termbase retrieval	number of lookup queries	0.41	0.44	0.47	<0.05
Terminology matches	number of correctly applied terms	0.56	0.59	0.62	<0.01

5.4 Contribution of Termbase Usage to Translation Performance

Hypothesis 3 (H3) was also supported: both the frequency of termbase retrievals and the number of terminology matches were positively correlated with translation quality. Table 5 reports the results.

As shown in Table 5, both termbase retrievals and terminology matches were positively correlated with translation performance, with stronger effects for accurate terminology use. This indicates that it was not retrieval alone but the effective application of terms that enhanced lexical accuracy and semantic adequacy. Taken together, these results validate H3 and highlight the importance of terminology management in translation learning, underscoring that consistent and correct use of retrieved terms is more impactful than mere consultation. A visual summary of the correlations between platform usage behaviors and translation performance is provided in Figure 6.

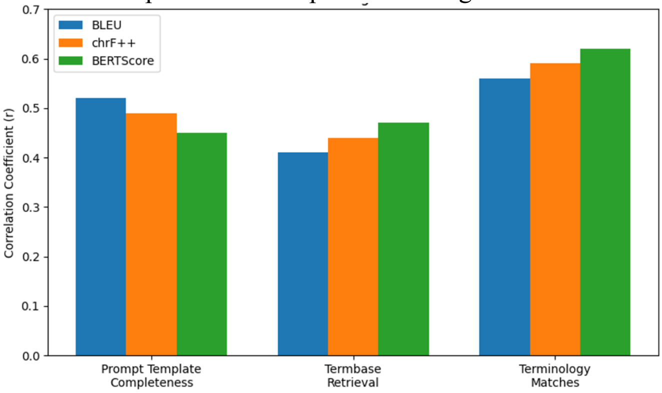

Fig. 6. Correlations Between Platform Usage Indicators and Translation Quality Metrics

6 Conclusions

6.1 Key Findings

This study developed a customized Chinese traditional culture translation platform “Sinolingo” (based on DeepSeek LLM) and validated its efficacy. Through paired t-test analysis of Task1 and Task2 data (n=30), the platform group outperformed the traditional tools group in BLEU, chrF++, and BERTScore ($p<0.05$, Cohen’s $d = 0.62$–0.78), validating Hypothesis 1 by addressing traditional learning issues and compensating for general-purpose LLMs’ cultural translation shortcomings. Pearson correlation analysis between Task 3 performance and platform usage logs indicated that prompt template completeness was significantly and positively associated with translation quality (BLEU: $r = 0.52$, chrF++: $r = 0.49$, BERTScore: $r = 0.45$, $p < 0.05$). In addition, termbase retrieval frequency showed moderate positive correlations with translation performance (BLEU: $r = 0.41$, chrF++: $r = 0.44$, BERTScore: $r = 0.47$, $p < 0.05$), while terminology matches exhibited stronger associations (BLEU: $r = 0.56$, chrF++: $r = 0.59$, BERTScore: $r = 0.62$, $p < 0.01$). These findings support Hypotheses H2 and H3, indicating that sustained and structured use of platform functions contributes to improved translation quality.

6.2 Resonance with Existing Research and Filling Research Gaps

This study aligns with Xu et al.’s view on AI platforms promoting translation reflection, quantifies feature usage-learning outcome correlations, solves general-purpose LLMs’ cultural translation inaccuracy via “self-built corpus + TRICE framework”[5]. Simultaneously, this study pioneers an exploratory approach by integrating LLMs with traditional cultural corpora within the realm of autonomous learning tools, aiming to provide novel insights for cultivating cultural translation professionals.

6.3 Educational Value

“Sinolingo” drives teaching innovation from rote knowledge indoctrination to competency development, establishes a tool design paradigm based on “functional modularization + embedded educational theory,” and focuses on Chinese cultural translation to cultivate cross-cultural talents meeting national needs.

6.4 Research Limitations and Future Prospects

While the present study yields encouraging results, its findings should be interpreted with appropriate caution. One primary constraint lies in the relatively limited sample size and the short duration of the experimental intervention, both of which inevitably restrict the generalizability of the conclusions. That said, post-hoc power analysis suggests that the statistical tests retained sufficient power, lending credibility to the observed effects and mitigating concerns regarding random variation.

6.5 Results on Culture-Loaded Translation Tasks

In addition to the quantitative analyses, a supplementary examination was conducted to explore the platform’s performance on culture-loaded translation tasks. The selected materials included culturally dense texts, such as classical Chinese writings, which involve implicit meanings, condensed imagery, and culturally embedded concepts.

A qualitative comparison between baseline translations and outputs generated using the Sinolingo platform revealed clear differences in the handling of cultural information. Baseline translations tended to simplify or flatten culturally salient elements, whereas translations produced under the TRICE framework more consistently preserved imagery and conveyed implicit meanings with greater coherence. These differences were particularly evident in metaphorical expressions and culture-specific references, where prompt constraints and corpus-based support facilitated more context-sensitive rendering. Although this analysis was qualitative in nature, it provides complementary evidence that the platform is well suited for translation tasks involving high cultural density.

These findings suggest that the advantages observed in quantitative metrics are also reflected in more interpretive translation scenarios, reinforcing the platform's applicability beyond general informational texts.

Acknowledgment

This study was funded by the Humanities and Social Sciences Program of the Ministry of Education (24YJA740030) and the research grant (LAI202303) from the Laboratory of Language and Artificial Intelligence, Guangdong University of Foreign Studies. It is also funded by National-level Undergraduate Innovative Training Program: Construction of a customized translation platform for traditional Chinese culture based on DeepSeek (202511846013).

References

1. Nguyen, T.N.N., Tran, T.T., Nguyen, N.H.A., Lam, H.P., Nguyen, H.M.S., Tran, N.A.T.: The benefits and challenges of AI translation tools in translation education at the tertiary level: A systematic review. International Journal of TESOL & Education, **5**(2), 132–148 (2025). https://doi.org/10.54855/ijte.25527
2. Yang, J. (2025). The integration of AI in translation pedagogy (PDF). David Publisher. Retrieved from https://www.davidpublisher.com/Public/uploads/Contribute/680ae8f3884e3.pdf
3. Xu, S., Sun, Y., Su, Y., & Liu, K. (2024). Translation pedagogy in the AI era: Design and application of a ChatGPT-powered platform for training and feedback. In S. Sun, K. Liu, & R. Moratto (Eds.), Translation Studies in the Age of Artificial Intelligence. Routledge. Archived at https://www.researchgate.net/publication/384968741
4. AlTwjiri, L., & Alghizzi, T. M. (2024). Investigating the integration of artificial intelligence in English as foreign language classes for enhancing learners' affective factors: A systematic review. Heliyon, **10**(10), e31053. https://doi.org/10.1016/j.heliyon.2024.e31053
5. Xu, S., Su, Y., & Liu, K. (2024). Integrating AI for enhanced feedback in translation revision: A mixed-methods investigation of student engagement. arXiv preprint arXiv:2410.08581. https://arxiv.org/abs/2410.08581
6. Wang, Q. (2024). A CEA framework-based report on the E-C translation practice of "Introduction to artificial intelligence and generative artificial intelligence" (Excerpts) (Master's thesis). Beijing Foreign Studies University. https://doi.org/10.26962/d.cnki.gbjwu.2024.001071
7. Zhu, Y. (2025). Research on corpus optimization strategies for AI translation from a cross-cultural communication perspective. Caizhi, (24), 5–8.
8. Ali, M., & Mohsen, M. (2024). Artificial intelligence in academic translation: A comparative study of large language models and Google Translate. Journal of Translation Studies, **15**(2), 101–120.
9. Bai, H., & Lei, X. (2025). TMNMT: Enhancing cross-language transmission of classical Chinese literature with multimodal neural machine translation. Scientific Reports, **15**(1), 15432. https://doi.org/10.1038/s41598-025-05921-1

10. Guo, H., Zhao, F., Cao, S., Lyu, X., Liu, Z., Wang, Y., Wang, B., Li, Z., Lu, C., Xu, Z., … & Hu, Y. (2025). Redefining machine translation on social network services with large language models. arXiv preprint arXiv:2504.07901.
11. Peng, K., Ding, L., Zhong, Q., Shen, L., Liu, X., Zhang, M., Ouyang, Y., & Tao, D. (2023). Towards making the most of ChatGPT for machine translation. Journal of Machine Translation, **18**(1), 45–66.
12. Sawyer, R.K. (ed.): The Cambridge Handbook of the Learning Sciences. Cambridge University Press, Cambridge (2006)
13. Kirschner, P.A., Sweller, J., Clark, R.E.: Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching.Educational Psychologist **41**(2), 75–86 (2006). https://doi.org/10.1207/s15326985ep4102_1
14. Zimmerman, B.J.: Attaining self-regulation: A social cognitive perspective. In: Boekaerts, M., Pintrich, P.R., Zeidner, M. (eds.) Handbook of Self-Regulation, pp. 13–39. Academic Press, San Diego (2000)
15. Pintrich, P.R.: A conceptual framework for assessing motivation and self-regulated learning in college students.Educational Psychology Review **16**(4), 385–407 (2004). https://doi.org/10.1007/s10648-004-0006-x
16. Winne, P.H., Hadwin, A.F.: Studying as self-regulated learning. In: Hacker, D.J., Dunlosky, J., Graesser, A. (eds.) Metacognition in Educational Theory and Practice, pp. 277–304. Lawrence Erlbaum Associates, Mahwah (1998)
17. Greene, J.A., Azevedo, R.: A theoretical review of Winne and Hadwin's model of self-regulated learning: New perspectives and directions. Review of Educational Research **77**(3), 334–372 (2007). https://doi.org/10.3102/003465430303953
18. Zimmerman, B.J.: Becoming a self-regulated learner: An overview. Theory into Practice **41**(2), 64–70 (2002). https://doi.org/10.1207/s15430421tip4102_2
19. Schunk, D.H., DiBenedetto, M.K.: Motivation and social-emotional learning: Theory, research, and practice.Contemporary Educational Psychology **60**,101830 (2020). https://doi.org/10.1016/j.cedpsych.2019.101832
20. Bowker, L., Fisher, D.: Computer-aided translation. In: Gambier, Y., van Doorslaer, L. (eds.) Handbook of Translation Studies, vol. 1, pp. 60–65. John Benjamins, Amsterdam (2010)
21. Wright, S.E., Budin, G. (eds.): Handbook of Terminology Management, vol. 2. John Benjamins, Amsterdam (2001)
22. Oxford, R.L.: Teaching and Researching Language Learning Strategies. Pearson Education, Harlow (2011)
23. Wenden, A.: Metacognitive knowledge and language learning. Applied Linguistics **19**(4), 515–537 (1998)