Effects of Translator–Computer Interaction Mode and Educational Levels on Textual Coherence Quality

Xinyuan Liu*, Xihua University
Weiqing Xiao, Shanghai International Studies University**
Sandra L. Halverson*, University of Agder**

ABSTRACT

This study explores the effects of three features of translator–computer interaction mode (translation resources, text segmentation and task roles), together with student educational levels, on textual coherence quality in Chinese-to-English translations. Fifty-two students at bachelor's and master's levels were divided into 26 pairs to complete four translation assignments in a classroom setting. Of the total, 35 submitted individual reflections on the challenges of textual problems in the assignments. The translations were coded for accuracy, completeness, coherence, normativity and appropriateness. The paper discusses overall translation quality and error counts along the five theoretical dimensions and then utilises generalised linear mixed-effects models on textual coherence quality. The findings indicate that educational levels have a significant main effect on coherence error presence; translation resources and task roles have main effects on coherence error frequency. A significant interaction effect between translation resources and task roles is observed on coherence error presence. Educational levels have a significant interaction effect with text segmentation on coherence error severity. The follow-up qualitative results support these findings that students' ability to detect coherence errors varies due to their different strategies to address coherence problems.

KEYWORDS

Translator-computer interaction mode, translation resources, text segmentation, task roles, educational levels, coherence quality assessment, interaction effects, L2 translation.

1. Introduction

Technological developments have increasingly complicated the interactions between human translators and technology, with profound impacts on translation quality. Accordingly, translation is regarded as human-computer interaction (O'Brien, 2012; Läubli & Green, 2019), translator-computer/technology interaction (Bundgaard et al., 2016; Pietrzak & Kornacki, 2021), or mediation between technologies with human translators (Moorkens, 2025). Technology-related resources such as translation memory (TM), term banks, machine translation (MT, including statistical or neural machine translation, i.e., SMT or NMT), and generative artificial intelligence (GenAI) have gradually been integrated into computer-assisted translation (CAT) tools for commercial translation workflows. Empirical evidence also shows that they have arguably lowered the task difficulty of post-editing of MT (MTPE) and improved overall translation quality for both specialised genres (Jia, 2024; Jia & Zheng, 2022) and literary tasks (e.g., Vieira et al., 2023).

Nevertheless, early assumptions about the "machine's failures in cohesion and coherence" (O'Brien, 2012, pp. 16-17) have persisted in recent empirical studies, while sentence segmentation in CAT tools remains an irritating function for translators trying to deal with text-level consistency and cohesion (Bundgaard et al., 2016; O'Brien et

* ORCID 0000-0003-3046-5944, e-mail: xinyuan712@hotmail.com
** ORCID 0000-0001-5954-5011, e-mail: wqxiao@shisu.edu.cn
*** ORCID 0000-0002-7078-5718, e-mail (corresponding author): sandra.l.halverson@uia.no

https://doi.org/10.26034/cm.jostrans.2026.8284

al., 2017). Translation pedagogy studies reveal that students' translation competence (Dong & Lan, 2010; Pan et al., 2022) and translation experience (Qian et al., 2022; Baer & Bystrova-McIntyre, 2009) have also contributed to predicting students' ability to detect coherence problems. However, students' development of translation competence follows a dynamic, non-linear, and spiral path (Albir et al., 2020; Mu et al., 2024). Educational levels (Łoboda & Mastela, 2023) might be a more stable alternative for data analysis, yet empirical support is needed.

This empirical study is part of a series exploring the effects of translator-computer interaction (TCI) mode on textual coherence quality of target texts (TT). We choose L1 to L2 (Chinese-to-English, C-E) translation for two reasons: first, Chinese is a high-context language that relies more on covert coherence features (Peng et al., 2023) that sometimes imply counterintuitive meaning in various contexts than on explicit cohesion features. This is different from English, which largely requires the overt and easily detectable signals of discourse markers (Halliday & Hasan, 1976); second, in L1 to L2 translation, it is more cognitively challenging to produce better quality than it is in L2 to L1 translation, for both from-scratch translation (Wang et al., 2022) and MTPE (Sun et al., 2025).

We identify three independent variables to capture TCI mode, in addition to educational levels (Section 2) and three dependent variables to analyse textual coherence quality (Section 3). Details of the data collection and analysis are then introduced (Section 4). The quantitative and qualitative results are discussed in Section 5 and the implications of this study are outlined in Section 6.

2. Literature review

To classify the independent and confounding variables related to TCI within the research landscape of MT/MTPE, we employed the factors affecting task difficulty of MTPE discussed by Jia and Sun (2023, p. 952), based on Meshkati's (1988) cohesive model of cognitive load. These are 'intrinsic task-related factors' and 'cognitive abilities' of task takers.

Translation resources are the first essential task-related feature for consideration. Previous studies show that NMT suggestions perform better than TM (Sánchez-Gijón et al., 2019; Vieira et al., 2023). Post-editing of SMT (Guerberof, 2014) or NMT (Jia et al., 2019) could deliver quality comparable to that of from-scratch translation or even better quality in accuracy (Jia & Sun, 2023) and L1 translation directionality (Sun et al., 2025). However, NMT does not seem to be sensitive enough to context-related problems across sentences and the whole document (Qian et al., 2022; Jiang & Niu, 2022; Moorkens, 2025; Tezcan et al., 2019). The second task-related feature, text segmentation, has also been studied. It has been shown that CAT tools' default sentence segmentation can reinforce translators' attention to local units, such as lexis and grammar, and to the syntactic structures of STs and TTs (Qian et al., 2022). This local level of text segmentation has made it difficult, awkward and even irritating for professional translators to detect textual problems, such as cohesion and coherence (Moorkens et al., 2018) or to maintain style consistency in their translation products (Bundgaard et al., 2016; Frankenberg-Garcia, 2022). Vieira et al. (2023) found that ten experienced literary translators generally made fewer keystrokes and editing visits in a paragraph segmentation task than in a sentence-based one. Similarly, Läubli et al.

https://doi.org/10.26034/cm.jostrans.2026.8284

(2022) reported that unsegmented presentation improved accuracy and efficiency when 20 professional translators revised anaphor errors between sentences. However, the two studies were limited by a small sample size (Vieira et al., 2023) or by using a research prototype instead of an available CAT tool (Läubli et al., 2022).

A third feature, task roles, is relatively underexplored as a task-related variable. From-scratch translators, post-editors and revisers (or reviewers, Guerberof, 2014) for monolingual or bilingual editing are common 'professional roles' (Carmo & Koponen, 2024) in translation workflows. Many studies have investigated translation quality in real-world CAT tools that enhance research ecological validity, instead of research-oriented platforms. However, they have often focused solely on (student or professional) translators' cognitive effort (Jia, 2024; Yao et al., 2025) and translation quality (Bundgaard et al., 2016) when using different translation resources, or when involved in different revision modes (Jia et al., 2019; Daems & Macken, 2020; Wang et al., 2024), such as editing of from-scratch translation and MTPE.

Higher levels of experience and translation competence lead to greater awareness of textual elements. This is supported by studies comparing translation quality or MTPE cognitive effort between students and professionals (e.g., Qian et al., 2022; Daems et al., 2017; Schaeffer, 2022), or across groups of professional translators (Bundgaard et al., 2016; Guerberof, 2014; Vieira et al., 2023). In studies of pedagogy experiments (Jia & Zheng, 2022; Jia, 2024; Wang et al., 2022; Wang et al., 2024; Yao et al., 2025) involving only students from the same educational level, students were reported to have a similar level of L2 proficiency but limited professional translation experience. In other words, these variables are often controlled.

Based on existing studies, three research gaps are apparent. First, limited studies have explored the effects of multiple task-related features of TCI and educational levels in work undertaken within CAT environments. Second, most research prioritises accuracy or acceptability of overall translation quality over textual coherence quality. Third, only a few studies have discussed controlling ST coherence as part of ST complexity. Apart from studies of ST word count (Wang et al., 2024), readability (Dai & Liu, 2024; Daems et al., 2017; Jia & Sun, 2023) and syntactic complexity (Jia & Sun, 2023), ST coherence remains underexplored in MT/MTPE empirical research, as a source of translation task difficulty shared by student translators and MT (Qian et al., 2022; Jiang & Niu, 2022). To our knowledge, one study (Yao et al., 2025) has measured ST coherence on a five-point rating scale for the English-to-Chinese translation direction, focusing on AI-assisted post-editing cognitive effort.

To fill these gaps, three key features characterise what we term 'TCI mode' (definitions in Table 2). They involve working procedures in which human translators, 1) utilise different kinds of *translation resources,* and 2) work with functions of *text segmentation* provided by the translation technology, as they 3) take on different *task roles* as translators, post-editors, or bilingual revisers. We also explore the effect of educational levels as a factor related to translator trainees' profiles, while translation experience and competence are also discussed as covariates.

https://doi.org/10.26034/cm.jostrans.2026.8284

3. Assessing textual coherence quality

Textual coherence is a crucial discourse feature in translation quality assessment (Feng & Yan, 2020), manifesting "across various error types" of quality such as accuracy, linguistic conventions, style and terminology (International Organisation for Standardisation, 2024, p. 16). Different from the metric of 'fluency' in some quality assessment frameworks (e.g., Silva et al., 2024; Zhang et al., 2024), textual coherence refers to "the way a text makes sense to the readers through the organization of its content, and the relevance and clarity of its concepts and ideas" (Richards & Schmidt, 2011, pp. 93-94), and may be achieved through overt grammatical, lexical or topical cohesion devices (Cui et al., 2022; Halliday & Hasan, 1976; Peng et al., 2023), or more covert features regarding discourse structures (Károly, 2017; Yuan, 2022). Previous research has primarily investigated overt cohesion features in translated texts, finding that they affect translation quality and reflect levels of translation competence (Dong & Lan, 2010; Pan et al., 2022). Tezcan et al. (2019, p. 42) added a new category of 'coherence', with seven sub-categories (i.e., logical problem, non-existing words, discourse marker, co-reference, inconsistency, verb tense), when assessing the 'fluency' aspect of NMT quality for English-to-Dutch literary translation.

To assess student translations, we employed five quality aspects proposed by the written translation competence scales in *China's Standards of English Ability* (Feng & Yan, 2020, pp. 53–58, Table 1), which offered detailed quality descriptors for teachers and students. Accuracy, completeness, coherence, normativity and appropriateness are regarded as five typical competence features of C–E translation (State Language Commission, 2024) and are predictors of overall translation quality (Lyu & Feng, 2024).

Aspects	Definitions	Examples from this study
Accuracy	A TT is lexically and grammatically error-free in expressing the content of its ST.	ST: 消息传开…… TT: As the news **spreaded** out Error type: grammatical typo Reference translation: As the news **spread**
Completeness	A TT conveys complete meaning in terms of content and information of its ST at the paragraph or text levels.	ST: 人们发现了**盛唐**流行的作画技法。 TT: People found the popular painting technique in the **Tang Dynasty**. Error type: incomplete content Reference translation: People found the painting technique popular during **the flourishing Tang Dynasty (c. 713-756 CE)**.
Coherence	A TT employs necessary cohesion devices to represent logical and functional relations in its ST.	ST: 1400多年前，隋炀帝在张掖建造了行宫，并主持召开了一次规模盛大的"万国博览会"。抵达张掖后，高昌、伊吾吐屯设等几十个国家的首领或者使臣前来**谒见**，表示愿意臣服。炀帝为此在附近草原上举行了长达6天的盛会。 TT (for the underlined): Yiwu and Tutunsad came to visit **them** and expressed their willingness to submit to the rule of the Emperor Yang of Sui. Error type: logical problem Reference translation: Upon arriving Zhangye, envoys from dozens of countries, such as Yiwu and Tutunsad, came to show their respects.
Normativity	A TT conforms to related standards or conventions of the target language in terms of numbers, proper nouns, measurement units, symbols, abbreviations, terms, sentence structures, and genre norms.	ST: 《隋书》 TT: **The Book of Sui** Error type: missing italics Reference translation: ***The Book of Sui***
Appropriateness	A TT exerts its sociolinguistic knowledge to satisfy the contextual requirements of its ST, including rhetorical devices, language varieties, and natural expressions.	ST: 盛陈珍宝文物丝绸锦缎。 TT: Silks and **jewelery** were put on display for the royalty. Error type: inconsistency of language varieties (British English required) Reference translation: Silks and **jewellery** were put on display for the royalty.

Table 1. Definitions and examples of C-E translation competence features

The overall translation quality assessment was based on weighted error counts, so previous error typologies for translated texts (Granger & Lefer, 2021) and NMT (Tezcan et al., 2019; Silva et al., 2024; Zhang et al., 2024) provided fine-grained

https://doi.org/10.26034/cm.jostrans.2026.8284

annotation guidelines for us to detect accuracy, coherence and normativity errors. Since preferential revision and changes also impact quality (Nitzke & Gros, 2020) in terms of accuracy and appropriateness, such changes by bilingual revisers on diction or style would not be considered erroneous if the translator's version is acceptable.

In addition to error frequencies along the five dimensions and the overall translation quality scores, three different exploratory measures of textual coherence quality were regarded as dependent variables (Table 2): (1) and (2) were straightforward error counts, and (3) was concerned with how problematic the textual coherence errors were.

Dependent variables	Levels	Criteria
(1) Coherence error presence	binary – yes/no	The two forms of frequency data were compared because of a methodological concern: which form, nominal or continuous, will be a better choice with a bigger effect size (if any)?
(2) Coherence error frequency	continuous from 0	
(3) Coherence error severity	ordinal and categorical • Error-free • Minor • Major • Critical	We explored the severity of coherence errors defined by error frequency bands. To avoid the impact of text length on each text but enhance replicability, we adopted statistically grounded thresholds. The severity level of Minor and Major was determined by a participant's coherence error frequency values exceeding + ¾ Standard Deviation (*SD*) from the Mean (*M*) value in every task for a specific role. Those over + ¾ *SD* from the doubled *M* value were classified as the Critical level.

Table 2. Dependent variables

As previous studies have implied, translation resources, text segmentation, task roles as task-related features, and educational levels as an individual feature will influence translators' tendency to overlook or focus on ST coherence features, further impacting the final quality. To avoid bias from examining only textual coherence quality, we first descriptively outlined all theoretical dimensions of overall translation quality (Section 5.1), then narrowed our focus to coherence itself (Sections 5.2 and 5.3). Our three research questions (RQ) are as follows:

RQ1: What is the level of student performance along five translation quality dimensions across different TCI mode features (translation resources, text segmentation, task roles) for both educational levels?
RQ2: How do translation resources, text segmentation and task roles affect coherence error presence, frequency and severity in students' translations?
RQ3: Do students' educational levels also affect coherence error presence, frequency and severity in students' translations?

4. Research design

4.1 The experimental set-up and hypotheses

As this study was integrated into a classroom setting, teaching ethics and equality were key concerns. To ensure that every student could experience different task roles, this TCI mode feature and educational levels were designed as within-subject independent variables (Table 3). Translation resources and text segmentation were treated as between-subject independent variables.

https://doi.org/10.26034/cm.jostrans.2026.8284

Independent variables		Working definitions	Levels	Tasks
TCI mode features	Translation resources	Translation resources are "all types of sources that translators consult to solve a translation problem" (Bundgaard & Christensen, 2019, p. 16). We examined two kinds of translation resources: TM and NMT. Any use of other translation resources that could impact translation quality, such as term banks, web engines, and GenAI, were provided or controlled under specific conditions (see section 4.3).	TM	M1, M2
			NMT	M3, M4
	Text segmentation	Text segmentation is the functional option of the selected "unit of text produced for a computer application to facilitate translation" (International Organisation for Standardisation, 2024, p. 3). We classify the units as sentence and paragraph segmentation.	Sentence	M1, M3
			Paragraph	M2, M4
	Task roles	Task roles refer to humans' professional positions in the language service and translation industry. We define a translator as a from-scratch human translator who "renders source language content into target language content in a written form" (International Organisation for Standardisation, 2024, p. 2). If they use any form of automatic translation technology for the draft translation, they could be regarded as post-editors. Bilingual revisers are those who examine "target language content against source language content for its suitability for agreed purpose" (International Organisation for Standardisation, 2024, p. 2). They revise and finalise the translation output from translators or post-editors. To have a balanced study design, we did not differentiate between translators and post-editors in the data modelling but discussed their differences in the results.	Translator	M1, M2 M3, M4
			Bilingual reviser	M1, M2 M3, M4
Educational levels		Educational levels are treated as an umbrella concept encompassing relevant aspects of translation experience and skill levels. The pedagogical rationale for selecting "educational level" is also grounded in the Chinese context of the translator education system (Association of Chinese Graduate Education, 2024; English Language Teaching Advisory Board under the Ministry of Education, 2020). The requirements of knowledge, competence, literacy, and professional ethics are implemented through a graduated scale throughout the bachelor's and master's programs of translation and interpreting. The levels thus demonstrate theoretical differences regarding learning outcomes and core courses such as Translation Technology.	Bachelor's	M1, M2 M3, M4
			Master's	M1, M2 M3, M4

Table 3. Independent variables

Based on previous research and focal variables in this study, we formulated three hypotheses for each RQ:

Hypothesis 1: In tasks with NMT, translators and bachelor's students will make more translation errors overall, leading to lower overall translation quality scores, than bilingual revisers and master's students.

Hypothesis 2: Translation resources, text segmentation and task roles will impact coherence error presence, frequency and severity. There will be more coherence errors with NMT, with sentence segmentation and with translators.

Hypothesis 3: Master's students will be more able to detect and categorise coherence errors, thus delivering fewer and less severe errors.

The four tasks (Table 3) were configured in a popular cloud-based service system for translation project management called Project Based Learning for Translation System (PBLT), developed by Sichuan Lan-bridge Information Technology Co., Ltd, China. We selected this platform because it offered a flexible option for text segmentation by sentence or paragraph, which helped operationalise the variable of text segmentation more efficiently than the options offered by other cloud-based CAT tools. For example, Figure 1 shows the PBLT interface for translators, and Figure 2 shows it for bilingual revisers. Their tasks were in blue.

https://doi.org/10.26034/cm.jostrans.2026.8284

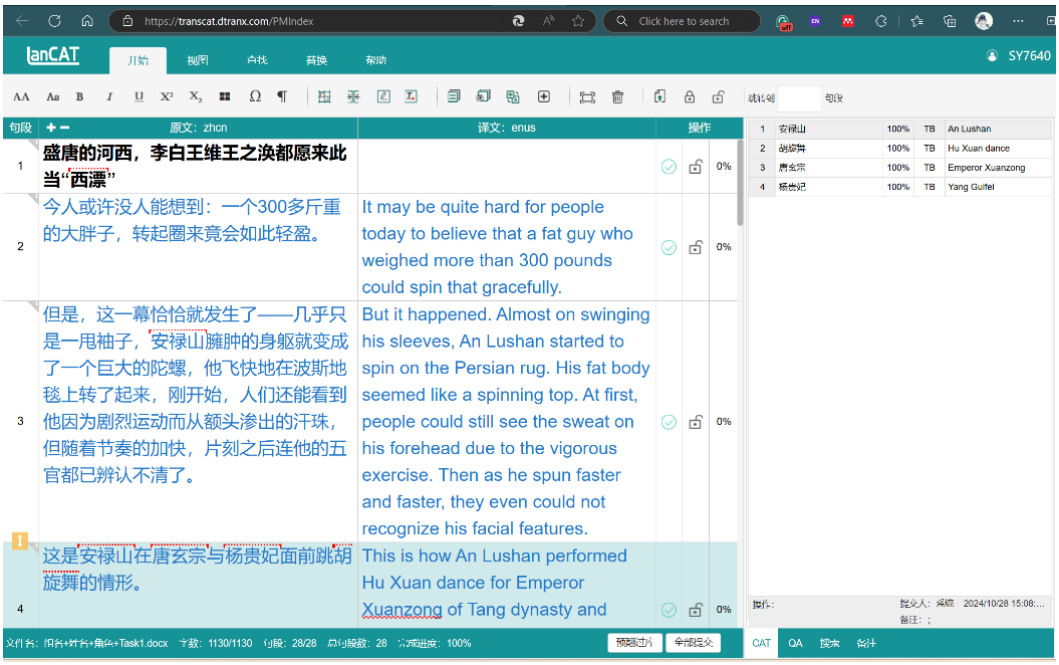

Figure 1. PBLT interface (screenshot) for translators (M1)

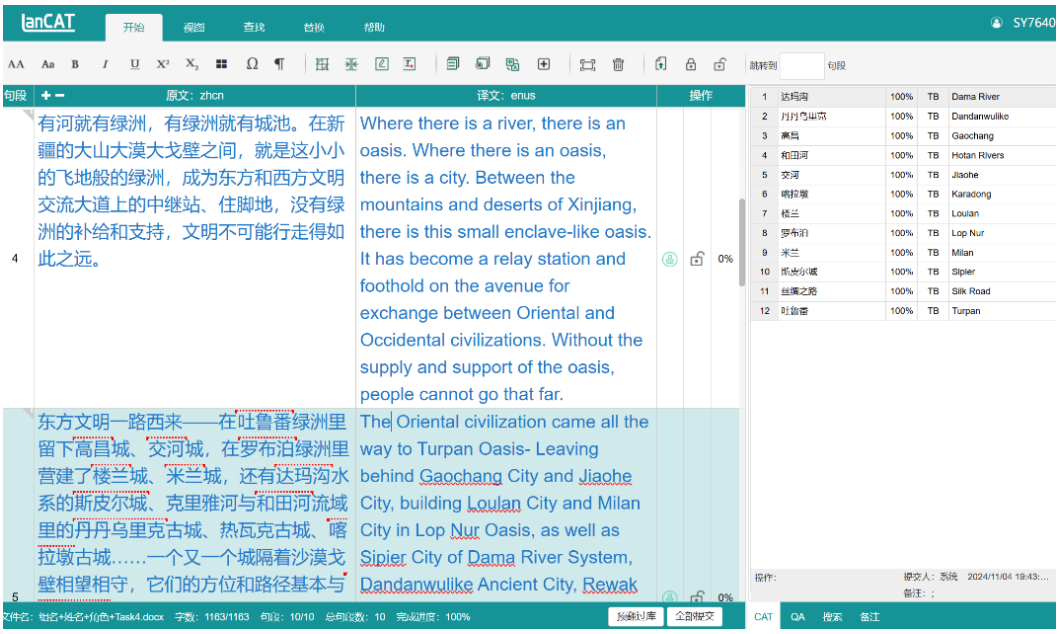

Figure 2. PBLT interface (screenshot) for bilingual revisers (M4)

Moreover, this company has developed its own integrated NMT engine, LanBridgeMT[1]. If the translation workflow is post-editing of NMT, NMT suggestions[2] will automatically appear in the TT areas. Students signed a translation affidavit for tasks M1 and M2, affirming that they would not seek assistance from NMT engines or GenAI tools. For M3 and M4, students could revise the draft English translation using LanBridgeMT and had full access to other NMT systems and GenAI tools, aiming to reduce reliance on a single NMT engine. In every task, students could use the term bank and web engines. We aligned C-E texts with 36,624 Chinese characters and 27,074 English words to create a TM in the PBLT platform. These texts were thematically similar to the research materials (Section 4.3) and were transcribed from high-quality publications, white papers, and the website of the online encyclopaedia *About China*.

https://doi.org/10.26034/cm.jostrans.2026.8284

We collected a term bank of 67 pairs with the help of two experienced professional translators.

4.2 Participants and procedures

A total of 16 female master's (M_{age}=23.88, *SD*=2.25) and 36 bachelor's students (N_{male}=3, N_{female}=33) aged from 19 to 23 years old (M_{age}=21.29, *SD*=1.99) from translation/interpreting or linguistics programmes participated in this study. Their first language was Chinese, and their second language was English. They all passed the Test for English Majors at Grade 4, with 92% having scored above 70. Eighteen students could be regarded as professionals (Mu et al., 2024) as they had passed the China Accreditation Test for Translators and Interpreters (CATTI)[3]; seven master's students possessed intermediate CATTI Level 2 and/or 3, and five master's and six bachelor's students had elementary CATTI Level 3.

All students completed the *Translation and Interpreting Competence Questionnaire* (Schaeffer et al., 2020), providing their self-reported overall translation competence (Table 4). Another online questionnaire enquired about students' learning experience with translation technology and their perception of task roles in translation projects. The aim of the self-perceived roles inferred from the responses was only for pre-experiment training. We used descriptions of the responsibilities of translators (including post-editors) and bilingual revisers from two Chinese national standards about translation services (China's Language and Terminology Committee, 2021, pp. 4–5; 2022, p. 9). All students regularly used GenAI tools or NMT, and fourteen students (N_B=8, N_M=6) knew little about post-editing and were unsure about the responsibilities of different task roles.

Item	Bachelor's (B)		Master's (M)	
	M	*SD*	*M*	*SD*
L1 to L2 translation competence*	68.61	6.69	76.44	13.92
L2 to L1 translation competence*	74.39	6.37	77.50	14.04
Weekly engagement of L1 to L2 translation (hour)	4.36	4.06	4.06	3.19
Weekly engagement of L2 to L1 translation (hour)	4.79	5.88	5.19	4.22
Years of translation experience	0.80	0.95	1.69	1.15

*: Students rated on a scale from 1 (very poor) to 100 (very good).

Table 4. Self-reported translation competence

From November to December 2024, students were divided into 26 pairs.[4] Each student took on the role of either a translator or a bilingual reviser and completed four sequential translation assignments in a course module for four weeks. Students were paired by educational levels. Bilingual revisers had more years of translation experience. The final valid number of participants was 48, comprising seven pairs of master's students and 17 pairs of bachelor's students. Since some students had less familiarity with post-editing workflows and task roles in the translation projects, we provided a two-hour practical session with professional technical support from the company before the assignments, along with a detailed translation brief for each task.

https://doi.org/10.26034/cm.jostrans.2026.8284

4.3 Materials

To ensure that research materials were comparable in terms of task difficulty, we collected both objective and subjective assessment data. First, we chose four coherent texts ($M_{character}$=439.75, *SD*=13.50) about the Hexi Corridor and ancient cities in China. We used the Chinese Text Analysis Platform (CTAP, Cui et al., 2022) and AlphaReadabilityChinese 2.0 (ARC 2.0, Lei et al., 2024; Lei & Zhang, 2025) to objectively compare ST complexity (Jia & Sun, 2023) and particularly cohesion complexity indicators. Both tools are fit for long-text analysis with complementary indicators. CTAP has 23 indicators that measure lexical, grammatical and topical cohesion within sentences and across paragraphs. ARC 2.0 aims to assess nine indicators of semantic complexity and one cohesion feature that CTAP lacks. Both CTAP and ARC 2.0 results showed that the four texts were similar in terms of ST complexity across Chinese characters, lexis, syntax, semantics and cohesion. Next, we invited four experienced C-E translation teachers ($M_{experience}$=16.25 years, *SD*=3.34) to assess overall translation difficulty using a five-point online questionnaire (Wang et al., 2022) on readability (*M*=2.75, *SD*=0.69), comprehensibility (*M*=2.75, *SD*=0.73) and translatability (*M*=3.75, *SD*=0.50). To avoid its effect, NMT quality (Jia & Sun, 2023; Jia & Zheng, 2022) was controlled at a similar level by assessment on a five-point scale, including adequacy (*M*=3.53, *SD*=0.86), acceptability (*M*=3.39, *SD*=0.92) and post-editing difficulty (*M*=2.65, *SD*=1.05). The two types of assessment data showed that the task difficulty was comparable at an above-intermediate level.

Tasks	ST	Character count	Lan-Bridge MT errors				
			Accuracy	Completeness	Coherence	Normativity	Appropriateness
M1	ST1	439	NA	NA	NA	NA	NA
M2	ST3	462	NA	NA	NA	NA	NA
M3	ST2	428	2	0	2	1	1
M4	ST4	430	2	0	2	1	1

NA: not applicable

Table 5. Task description

For each task, we identified up to five potential coherence errors that would be difficult for students to detect, based on our teaching experience. The texts were randomly assigned to each task. For M3 and M4 (Table 5), the five types of errors for texts two and four, respectively, were naturally generated by NMT (i.e., Lan-BridgeMT) without any manipulation. Such a measure avoided the confounding effect of an imbalance in the number of NMT errors on coherence error severity.

4.4 Data analysis

The five types of translation errors were annotated according to the framework in Table 1, and the overall translation quality scores were then calculated (Figure 3). Researcher A was the class teacher and checked the final scoring results, and researcher B worked as the teaching assistant. As she was familiar with the written translation competence scales and language testing, researcher B was responsible for annotating errors and scoring students' performance and repeated the assessment process one month later. Each normativity error was penalised by 0.25 points, and the remaining four by one point, respectively. Every student's performance on the five

https://doi.org/10.26034/cm.jostrans.2026.8284

types of errors was compared with the average error counts for each text and task role. Finally, we scored the overall translation quality on a scale of 1 (worst) to 6 (best). The repeated intra-rater Kappa value reached a reliable level of 0.87. Finally, a total of 1,764 valid observations of quality data were analysed.

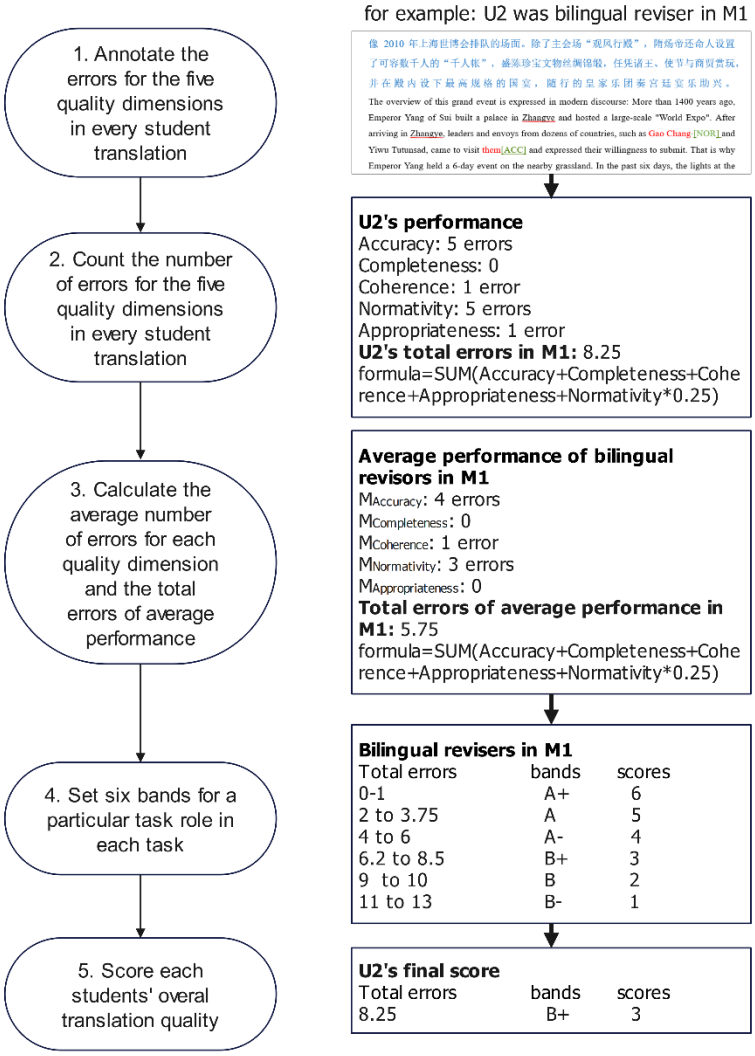

Figure 3. Assessment process

As participants were nested within task roles, we included random intercepts and random slopes for repeated measurements (2-4 observations per participant, 48-50 observations per item) across different task roles. The five experience-related variables (all three levels: low, intermediate, and high)[5] in Table 3 and professional certification (three levels: novice, elementary, and intermediate) were treated as covariates, because individual variation in these aspects could exist within the same educational level. Categorical fixed effects and covariates were coded using successive differences contrasts, implemented by the contr.sdif() function from the MASS package (Venables & Ripley, 2002), which compared each level with the preceding one.

We fitted the generalised linear mixed-effects models using the glmer() function from the lme4 package (Bates et al., 2015) for coherence error presence. A binomial

https://doi.org/10.26034/cm.jostrans.2026.8284

distribution with a logit link was specified. For coherence error frequency, we compared the observed proportion of zeros (17.3%) with the expected proportion under a Poisson distribution (20.8%), indicating no zero inflation. Furthermore, the count outcome was underdispersed, as its variance (1.31) was less than its mean (1.57). We finally fitted Conway-Maxwell Poisson (COMP) mixed-effects models (log link) using the glmmTMB (Brooks et al., 2017) package, as they significantly outperformed the standard Poisson models. We fitted the cumulative link mixed models using the clmm() function from the ordinal package (Christensen, 2025), specifying a logit link, for coherence error severity. Diagnostic checks of the final models included multicollinearity using the check_collinearity() function and post-hoc comparisons using the emmeans() and contrast() functions.

To qualitatively assess RQ2 and RQ3, we requested student feedback on textual translation problems (instead of textual coherence) to avoid participants' potential speculation of research aims. Finally, three master's and 32 bachelor's students submitted video reflections in response to the three questions in Figure 4.

Dear students:
1. Please watch the 50-minute video feedback on your group performance on the four assignments from teachers.
2. Answer the following three questions, and upload a five-minute video reflection to the Blackboard Platform. DDL: December 31st, 2024.

Q1: How did you individually or collaboratively solve translation problems at the textual level? Please share a case.
Q2: In which assignment(s) did you find it easier to detect translation problems at the textual level? Please give your reasons.
Q3: In which assignment(s) did you find it harder to detect translation problems at the textual level? Please give your reasons.

Figure 4. Questions for students' reflection

MAXQDA Analytics Pro 2020 was employed for content analysis of the transcription of 35 students' feedback on the framed questions, including 1,127 English words and 35,545 Chinese characters. Combining deductive and inductive coding formations (Kuckartz & Rädiker, 2019), researcher B repeated the coding process every one or two months from February to May 2025, to ensure inclusion of all relevant information. The first coding served to familiarise the content (952 codes) related to textual coherence problem-solving. We calculated the document-level intra-coder agreement (Kuckartz & Rädiker, 2019, p. 273) between the second (762 codes) and third (428 codes) coding results, with 93% matching codes for 'TCI mode' (94 codes), 94% for 'educational levels' (32 codes), 62% for 'strategies' (208 codes), 28% for 'difficult tasks', 45% for 'easy tasks' and 23% for 'other factors'. Finally, we developed a coding system that included only codes that reached 70% agreement[6] or higher (Cheung & Tai, 2023).

https://doi.org/10.26034/cm.jostrans.2026.8284

5. Results

5.1 Overall translation quality

We used the ggplot() function to illustrate the pattern of three features of TCI mode and educational levels on overall translation quality scores and error frequencies within the five dimensions. Given their dependence on translator drafts, bilingual revisers' outputs, including corrections or introductions of any type of errors, were treated as their final products. Bilingual revisers spotted and corrected more errors for each dimension (Figures 5 and 6), so that their versions were usually much better than translators' versions across the four tasks, particularly in M2 and M4. A lower median overall translation quality was observed among bachelor's students than among master's students in M4 when they played different task roles.

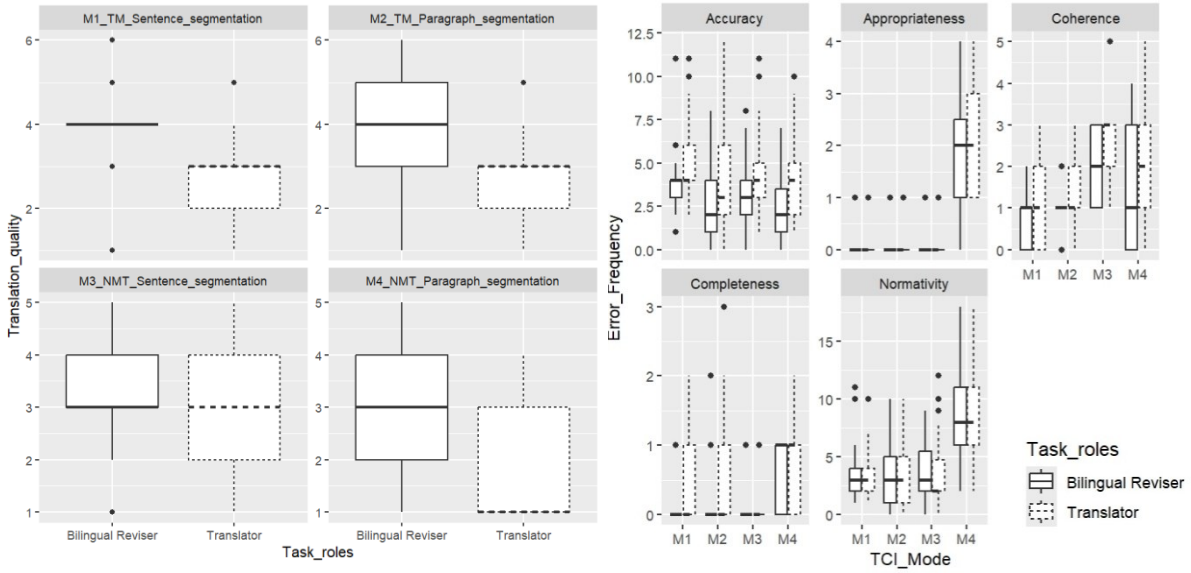

Figure 5. Boxplots of overall translation quality (N=196) and errors (N=980) by task roles

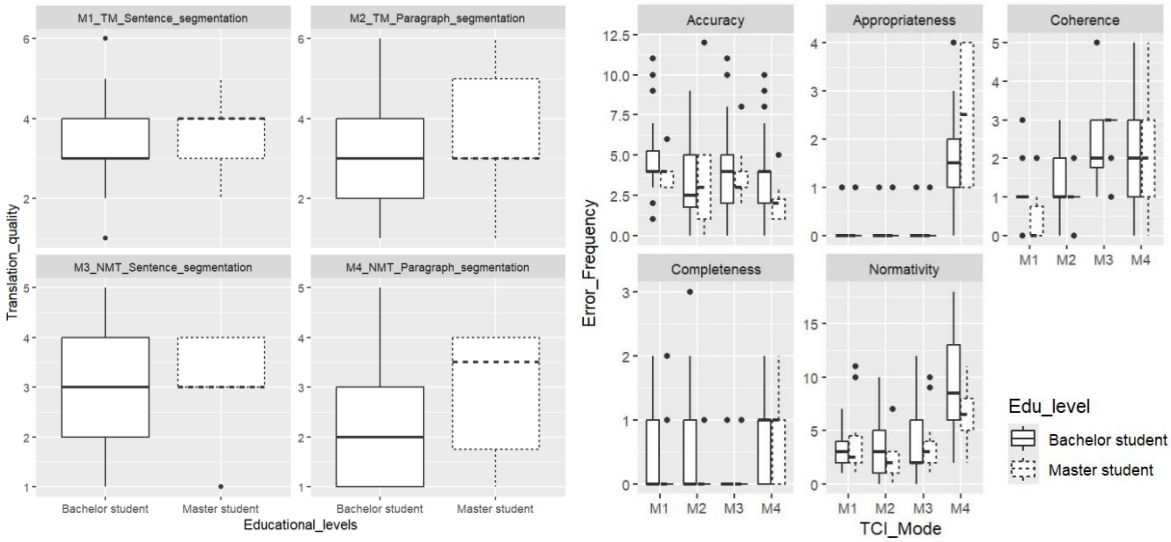

Figure 6. Boxplots of overall translation quality (N=196) and errors (N=980) by educational levels

https://doi.org/10.26034/cm.jostrans.2026.8284

It was unsurprising that there were fewer errors in accuracy and completeness made by bilingual revisers and master's students (Figure 6). However, in most cases, students in neither task roles could easily recognise appropriateness translation errors, as this required careful consultation of contextual expressions. Bachelor's students paid less attention to the adherence to norms in their final products, even though they were required to do so. But across different educational levels, master's students could correct many errors of normativity. Different distributions of coherence errors were also presented, which are the focus of the following sections on textual coherence quality, measured in terms of coherence error presence, frequency, and severity.

5.2 Textual coherence quality

5.2.1 Overview of models

In terms of textual coherence quality in particular, many translators could not detect coherence errors in NMT and introduced new ones due to their inadequate understanding of textual relations, as shown by more frequent and more severe coherence errors in the NMT conditions (Table 6) and in sentence segmentation. Bilingual revisers varied in how they addressed their partners' coherence errors across tasks.

Independent variables	Tasks	Coherence error presence		Coherence error frequency		Coherence error severity			
Task roles		Yes	No	M	SD	Error-free	Minor	Major	Critical
Translator	M1	17	8	1.08	0.98	8	10	4	3
	M2	21	4	1.44	0.99	4	21	0	0
	M3	26	0	2.50	0.89	0	11	14	1
	M4	24	1	2.24	1.34	1	20	2	2
Bilingual reviser	M1	16	9	0.72	0.60	9	14	2	0
	M2	20	4	1.04	0.61	4	20	0	0
	M3	23	0	2.00	0.78	0	7	9	7
	M4	15	8	1.52	1.35	8	7	8	0

Table 6. Textual coherence quality across four tasks (N=588)

The generalised variance inflation factors for independent variables and interaction effects were all less than 3.5 (Tables 7, 8, 9), indicating no multicollinearity. The null mixed-effects models of coherence error presence revealed substantial item-level variance but negligible between-participant variability. Therefore, we retained only Item random effects for parsimony. The final model of coherence error presence ($\chi^2(-5) = 15.21$, $p = .009$) was a significantly better fit than its null model.

https://doi.org/10.26034/cm.jostrans.2026.8284

	Parameters	Variance	SD	β	SE	z	p	odds ratio (OR)	95%CI	AIC	contrasts	Likelihood ratio tests		VIF
												χ^2(*df*=1)	*p*	
Null model	glmer(formula = Coherence_error_presence ~ 1 + (1 | Item))													
	Intercept	-	-	1.99	0.73	2.75	.006 ***	7.35	[1.72, 30.46]	172.41	-	-	-	-
	Random effect (Item)	1.68	1.30	-	-	-	-	-	-	-	-	-	-	-
Final model	glmer(formula = Coherence_error_presence ~ Translation_resources + Text_segmentation + Task_roles + Educational_levels + (1 | Item) + Translation_resources:Task_roles)													
	Random effect (Item)	1.17	1.08	-	-	-	-	-	-	-	-	-	-	-
	Fixed effects													
	Intercept	-	-	1.82	0.73	2.50	**.003** ***	8.32	[2.04, 33.86]	167.20	-	-	-	-
	Translation_resources2-1	-	-	-2.21	1.39	-1.59	.112	0.10	[0.01, 1.70]	-	1=NMT 2=TM	-	-	1.26
	Text_segmentation2-1	-	-	0.08	1.18	0.07	.903	1.16	[0.11, 12.22]	165.22	1=Paragraph 2=Sentence	0.02	.902	1.14
	Task_roles2-1	-	-	1.36	0.61	2.24	**.021** *	4.01	[1.23, 13.06]	-	1=Bilingual Reviser 2=Translator	-	-	1.89
	Educational_levels2-1	-	-	-1.23	0.46	-2.67	**.022** *	0.37	[0.16, 0.87]	170.33	1=Bachelor student 2=Master student	5.13	**.024** *	1.01
	Translation_resources2-1:Task_roles2-1	-	-	-2.36	1.21	-1.95	**.039** *	0.08	[0.01, 0.89]	170.73	-	5.53	**.019** *	1.89
Note: β=Estimate; SE=Standard Error; *z*=z-value; *p*=significance level (*p<0.05, ** p<0.01, *** p<0.001)														

Table 7. Null and final models of TCI modes and educational levels on coherence error presence

For the models of coherence error frequency, we retained only the Item random effects, while the Participant random effects showed minimal variance. The final model fit much better than its null model ($\chi^2(4) = 21.24$, $p < .001$).

	Parameters	Variance	SD	β	SE	z	p	Incidence rate ratio (IRR)	95%CI	AIC	contrasts	Likelihood ratio tests		VIF
												χ^2(*df*=1)	*p*	
Null model	glmmTMB(formula = Coherence_error_frequency~1+(1|Item))													
	Intercept	-	-	0.40	0.18	2.21	.027 *	1.49	[1.04, 2.12]	562.60	-	-	-	-
	Random effect (Item)	0.12	0.35	-	-	-	-	-	-	-	-	-	-	-
Final model	glmmTMB(formula = Coherence_error_frequency ~ Translation_resources + Text_segmentation + Task_roles + Educational_levels + (1 | Item))													
	Random effect (Item)	0.004	0.07	-	-	-	-	-	-	-	-	-	-	-
	Fixed effects													
	Intercept	-	-	0.35	0.07	5.17	**2.35e-07** ***	1.42	[1.24, 1.62]	549.36	-	-	-	-
	Translation_resources2-1	-	-	-0.64	0.12	-5.29	**1.24e-07** ***	0.53	[0.42, 0.67]	556.12	1=NMT 2=TM	8.76	**.003** **	1,04
	Text_segmentation2-1	-	-	-0.08	0.11	-0.74	.458	0.92	[0.74, 1.14]	547.92	1=Paragraph 2=Sentence	0.57	.452	1.05
	Task_roles2-1	-	-	0.32	0.09	3.43	**.0001** ***	1.38	[1.15, 1.65]	559.05	1=Bilingual Reviser 2=Translator	11.69	**.00063** ***	1.00
	Educational_levels2-1	-	-	-0.09	0.11	-0.85	.398	0.91	[0.74, 1.13]	548.09	1=Bachelor student 2=Master student	0.73	.393	1.02
Note: β=Estimate; SE=Standard Error; *z*=z-value; *p*=significance level (*p<0.05, ** p<0.01, *** p<0.001)														

Table 8. Null and final models of TCI modes and educational levels on coherence error frequency

For models of coherence error severity, the random intercept variances for Item and Participant were retained. The final model of coherence error severity provided a better fit than its null model ($\chi^2(8) = 35.06$, $p < .001$).

https://doi.org/10.26034/cm.jostrans.2026.8284

	Parameters	Variance	SD	*β*	SE	*z*	*p*	odds ratio (OR)	95%CI	AIC	contrasts	Likelihood ratio tests		VIF
												$\chi^2(df=1)$	*p*	
Null model	clmm(formula = Coherence_error_severity~1+(1| Participant)+(1| Item))													
	No coefficients	-	-	-	-	-	-	-	-	413.99	-	-	-	-
	Random effect (Item)	1.12	1.06	-	-	-	-	-	-	-	-	-	-	-
	Random effect (Participant)	0.18	0.42	-	-	-	-	-	-	-	-	-	-	-
Final model	clmm(formula = Coherence_error_severity ~ Translation_resources + Text_segmentation + Task_roles + Educational_levels + +Educational_levels:Translation_resources + Educational_levels:Text_segmentation + (L2translation_competence) + (1 | Participant) + (1 | Item))													
	Random effect (Item)	0.35	0.59	-	-	-	-	-	-	-	-	-	-	-
	Random effect (Participant)	0.18	0.42	-	-	-	-	-	-	-	-	-	-	-
	Fixed effects													
	Coefficients	-	-	-	-	-	-	-	-	394.93	-	-	-	-
	Translation_resources2-1	-	-	-2.06	0.72	-2.84	**.005 ****	0.13	[0.03, 0.53]	-	1=NMT 2=TM	-	-	3.31
	Text_segmentation2-1	-	-	-0.90	0.64	-1.40	.163	0.41	[0.12, 1.43]	-	1=Paragraph 2=Sentence	-	-	3.34
	Task_roles2-1	-	-	0.09	0.30	0.31	.759	1.10	[0.61, 1.97]	393.02	1=Bilingual Reviser 2=Translator	.09	.759	1.15
	Educational_levels2-1	-	-	-1.00	0.42	-2.37	**.018 ***	0.37	[0.16, 0.84]	-	1=Bachelor student 2=Master student	-	-	2.06
	L2translation_competence 2-1	-	-	-0.97	0.53	-1.85	.064	0.38	[0.14, 1.06]	398.57	1=High 2=Intermediate 3=Low	7.64	**.022 ***	1.51
	L2translation_competence 3-2	-	-	-1.17	0.59	-1.97	**.049 ***	0.31	[0.10, 0.99]					
	Translation_resources2-1:Educational_levels2-1	-	-	-1.32	0.72	-1.85	.065	0.27	[0.07, 1.09]	396.33	-	3.40	.065	1.20
	Text_segmentation2-1:Educational_levels2-1	-	-	-2.99	0.82	-3.64	**.0003 *****	0.05	[0.01, 0.25]	407.56	-	14.63	**.0001 *****	1.28
	Threshold coefficients													
	Error-free | Minor	-	-	-1.77	0.44	-4.02	-	0.17	[0.07, 0.40]	-	-	-	-	-
	Minor | Major	-	-	1.82	0.46	3.98	-	6.17	[2.52, 15.10]	-	-	-	-	-
	Major | Critical	-	-	3.94	0.57	6.91	-	51.65	[16.88, 158.08]	-	-	-	-	-

Note: *β*=Estimate; SE=Standard Error; *z*=z-value; *p*=significance level (**p*<0.05, ** *p*<0.01, *** *p*<0.001)

Table 9. Null and final models of TCI modes and educational levels on coherence error severity

5.2.2 Effects of TCI mode

The simple effects analysis indicated an interaction effect between translation resources and task roles on coherence error presence (Table 7). For the Translator group, a significantly higher probability of coherence errors (OR = 33.41, 95% CI [1.19, 935.9], SE = 56.80, $z = 2.06$, $p = .039$) was observed in NMT tasks than in TM tasks. In contrast, for the Bilingual Reviser group, students were more likely to make coherence errors using NMT than TM, but the difference was not significant (OR = 2.81, 95% CI [0.19, 42.3], SE = 3.89, $z = 0.75$, $p = .455$).

https://doi.org/10.26034/cm.jostrans.2026.8284

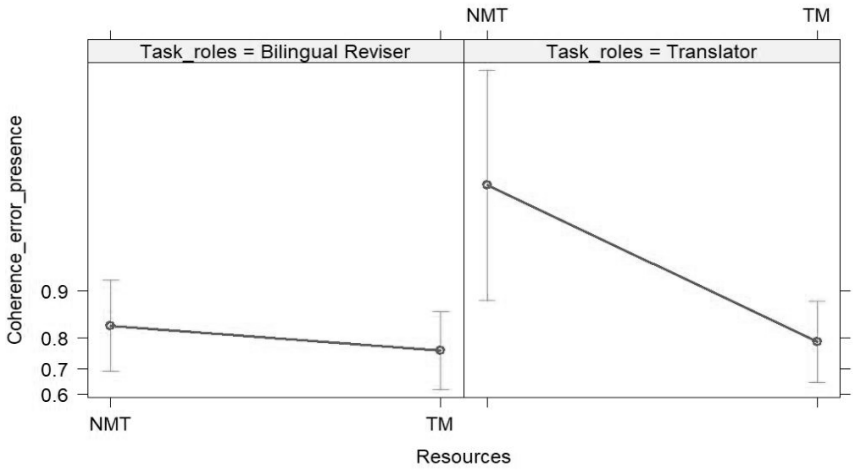

Figure 7. Interaction effect plot of translation resources and task roles on coherence error presence (N=196)

Only translation resources and task roles had significant main effects on coherence error counts (Table 8). When the translation resource was TM, it indicated a lower probability of making more coherence errors. Translators were more likely to have more frequent coherence errors than bilingual revisers. A significant main effect of translation resources was also found on coherence error severity (Table 9). Students produced significantly fewer severe coherence errors in tasks assisted by TM than in those assisted by NMT. Text segmentation did not have significant main effects on coherence error presence (Table 7), frequency (Table 8) or severity (Table 9).

5.2.3 Effects of educational levels

The simple effects analysis revealed a significant interaction effect between educational levels and text segmentation on coherence error severity (Figure 8). Master's students made significantly fewer severe coherence errors (OR = 10.89, 95%CI [1.85, 64.2], SE = 0.91, z = 2.64, p = .008) in the paragraph segmentation tasks than in sentence segmentation tasks. For bachelor's students, no significant difference was found between the sentence and paragraph segmentation conditions (OR = 0.55, 95%CI [0.18, 1.73], SE = 0.58, z = -1.02, p = .306).

https://doi.org/10.26034/cm.jostrans.2026.8284

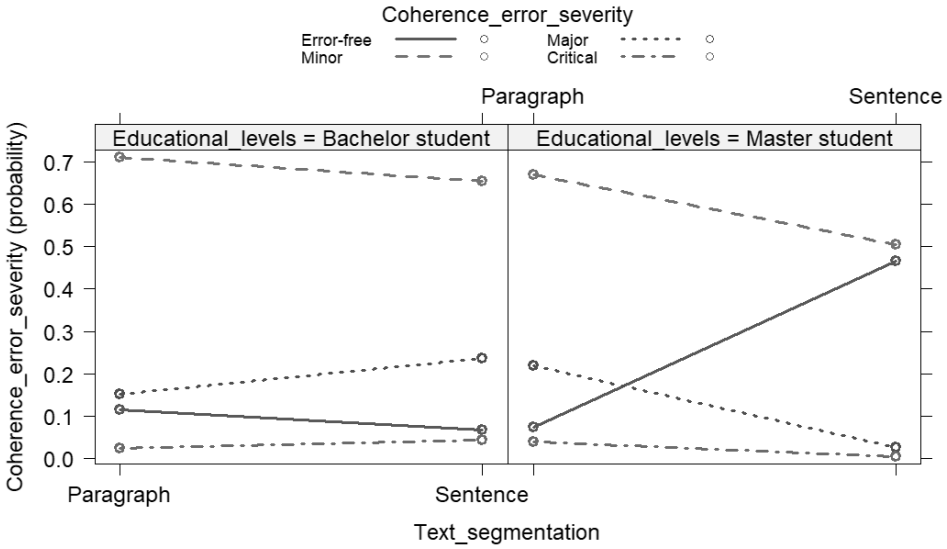

Figure 8. Interaction effect plot of text segmentation and educational levels on coherence error severity (N=196)

Educational levels had a significant main effect on coherence error presence (Table 7). Master's students were found to have a significantly lower probability of making coherence errors than bachelor's students. For all covariates, we found only that L2 translation competence had a significant main effect on coherence error severity (Table 9). Regardless of educational levels, students who reported low L2 translation competence unexpectedly had a lower probability of making severe coherence errors than their peers who evaluated their L2 translation competence as intermediate.

5.3 Challenges for textual coherence problems

Qualitative analysis of students' reflections (Figure 9) revealed a similar pattern of complex relationships among the three TCI mode features and educational levels regarding textual coherence quality.

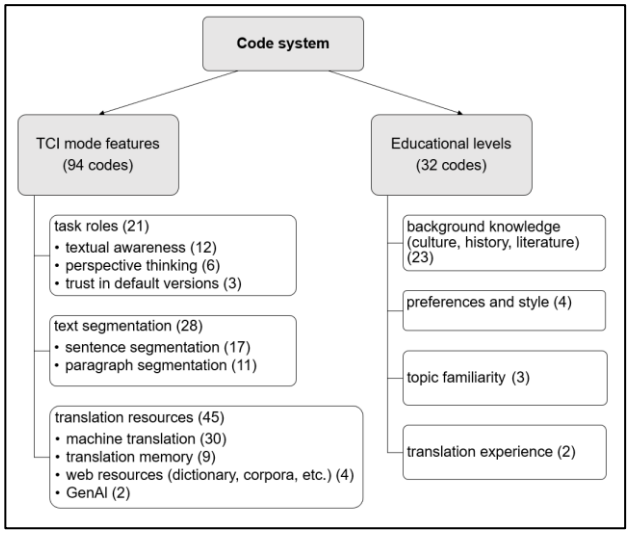

Figure 9. Final coding structure and analysis of student reflections[7]

The most challenging combination of TCI mode features for translators was NMT and sentence segmentation. For example (Figure 10), students reported:

https://doi.org/10.26034/cm.jostrans.2026.8284

Example 1

1) M8 (a master's student)
"Because in some tasks the system divided the STs into sentences, this visual change required much more attention to keep consistent semantic relations within sentences. Sometimes this also prevented me from taking very bold decisions to add or delete information in the STs."
[translated from Chinese]

2) B7 (a bachelor's student)
"I found the first two tasks (i.e., M1 and M3) more challenging in terms of finding logical order. Because when I'm doing these projects, I would just focus on the single sentence presented on my screen and then move to the next sentence without remembering the previous one."
[originally in English]

Figure 10. Example of students' reflection

Students expressed differing viewpoints on the effects of translation resources and text segmentation in addressing textual problems. For students familiar with the topics and who tended to analyse ST coherence features, M4 with NMT and paragraph segmentation was much easier because they could do simultaneous comparative reading between the ST and the TT. They were able to understand how information developed by observing semantic relations and logical relations across sentences and paragraphs. However, other students found tasks involving NMT more demanding, regardless of the type of text segmentation. They reported that it was difficult to analyse the main idea and logical relations across sentences efficiently because their judgements were influenced by syntactic structures and by seemingly fluent TTs produced by NMT.

Regarding the effect of task roles, translators tended to focus more on processing issues of sentence-level accuracy that they did not understand, so textual problems could not be prioritised. Many students pointed out that textual problems, particularly logical coherence, were much more obvious, and some would regard these problems as their revision focus when they were bilingual revisers.

Example 2
Task Four (M4: NMT + paragraph segmentation)
ST: 有河就有绿洲，有绿洲就有城池。在新疆的大山大漠大戈壁之间，就是这小小的飞地般的绿洲，成为东方和西方文明交流大道上的中继站、住脚地，没有绿洲的补给和支持，文明不可能行走得如此之远。
NMT (for the underlined sentence): […] [**there is**] this small enclave-like [**oasis**] […]
Error type: co-reference
Reference translation: […] these small patches of oases […]
B25 (translator): […] there exists [**this**] small, enclave-like [**oasis**] […]
B26 (bilingual reviser): […] there exist [**these**] small, enclave-like [**oases**] […]

Figure 11. Example of a coherence problem

In the example (Figure 11), the translator B25 did not realise that the NMT version was incorrect regarding the textual relation between the demonstrative pronoun *这* 'this or the or these' and *绿洲* 'oasis or oases' in the ST. B26 corrected B25's use of the singular form of *绿洲* 'oasis' to the plural, and commented that inference from the whole text suggested that there were many oases in the ST. This indicates that the bilingual revision phase allowed them to read the TTs from a broader perspective and from the English readers' perspective, and then to compare the logical relations of STs and TTs.

https://doi.org/10.26034/cm.jostrans.2026.8284

6. Discussion and future research directions

This empirical study explored the impact of three TCI mode features and educational levels on multidimensional overall translation quality and on textual coherence quality in students' C-E translation. Regarding the valid measurements of textual coherence quality assessment, coherence error presence, frequency and severity were found to be differently sensitive to variations in translation resources, text segmentation, task roles and educational levels.

First, Hypothesis 1 is confirmed. Our finding that translators indeed made more errors than bilingual revisers, because their attention and task focus were different, is consistent with previous studies (Baer & Bystrova-McIntyre, 2009; Pietrzak & Kornacki, 2021; Swar & Mohsen, 2023). The trend is similar for bachelor's and master's students. Accuracy errors were easily avoided, detected, and corrected by both task roles and educational levels, leaving many appropriateness, normativity, and coherence errors unaddressed. Students mentioned in their qualitative feedback that they lacked sufficient attention and time to process textual problems after addressing low-cognitive-level problems, such as spelling errors (Qian et al., 2022; Frankenberg-Garcia, 2022), words or phrases. Consequently, textual elements were neglected. Hypothesis 2 is partially confirmed. Translation resources and task roles, as two task-related TCI mode features, had main effects on coherence error frequency. This finding is consistent with previous research (Qian et al., 2022; Jiang & Niu, 2022; Moorkens, 2025; Tezcan et al., 2019; Swar & Mohsen, 2023) that MTPE will result in more coherence errors when human translators are unaware of those introduced by NMT. Our new finding of a significant interaction effect on coherence error presence between translation resources and task roles further reveals that task roles matter a lot in dealing with textual elements.

Hypothesis 3 is partially confirmed, as educational levels had a main effect on coherence error presence. This provides new empirical support for Łoboda & Mastela's (2023, p. 520) suggestion that the master's level be the "earliest advisable stage" for task role training, as they could "better identify and categorise errors". The coding theme on 'strategies', although with relatively low agreement, offered some valuable hints. The three master's and two outstanding bachelor's students reported that they had accumulated background knowledge of cultures, histories, and literature related to the topics, particularly through systematic learning experiences in translation projects, enabling them to develop diverse coping strategies. However, a more balanced distribution of participants across educational levels, along with their qualitative feedback, is necessary to explain the impact of educational levels.

Significant interaction effects on coherence error severity were observed for educational levels with text segmentation. Although our qualitative analysis supported the findings of Vieira et al. (2023) and Läubli et al. (2022) that paragraph segmentation in CAT tools reduced cognitive demand for translators, this condition surprisingly led to different distributions of coherence error severity across both educational levels. An unexpected positive effect was found in ensuring less coherence severity for master's students in the sentence segmentation tasks, while bachelor's students had a higher probability of making major and critical coherence errors in the sentence segmentation tasks than in the paragraph segmentation tasks. This was probably due to prior use of CAT tools. Of the total, 71% of the master's students were familiar with the sentence

https://doi.org/10.26034/cm.jostrans.2026.8284

segmentation translation tasks in the CAT tools, while 86% of the bachelor's students usually completed their tasks within the Microsoft Word documents.

L2 translation competence played a role in explaining the coherence error severity, though with a pattern different from the positive effect of translation competence (Dong & Lan, 2010; Pan et al., 2022; Schaeffer, 2022). Students with higher L2 translation competence may have underestimated their scores. We did not observe any effects from translation experience or professional certification, which does not align with the effect of translation experience found in previous studies (Guerberof, 2014; Daems et al., 2017). Such experience-related information from students might require more fine-grained measurements, as students typically accumulate translation experience through classroom projects, homework, or limited internship activities.

This study has addressed the outlined research gaps, but future investigations should consider the following methodological aspects to address our research limitations. First, the task roles in our study were based solely on typical linguistic roles in the language service industry workflow. It would be interesting to investigate how a range of task roles or stakeholders respond to coherence problems. Second, the generalisability of educational levels might be limited within Chinese educational contexts. Future research could use well-designed L2 proficiency tests (e.g., Chung, 2020) to examine the potential effects of L2 proficiency on textual coherence quality across different student groups. Third, this study also found that the Item random effects (i.e., texts) were the primary source of variability, indicating additional explanatory power for textual coherence quality. Besides, many students noted that the four texts exhibited different textual coherence features, particularly functional relations (Mann and Thompson, 1987) within and across sentences, as well as various levels of covert logical structures. The extent to which ST coherence features affect TT textual coherence quality warrants further exploration. Finally, although multilevel generalised regression models would offer greater robustness[8], our study is likely to suffer from less stable modelling results due to a limited sample size and the exploratory nature of the dependent variables. Future studies would benefit from multilevel structural equation modelling to explore causal pathways involving educational levels, strategies, and textual coherence quality, using larger sample sizes.

Acknowledgements

This study was ethically approved by Shanghai Key Laboratory of Brain-Machine Intelligence for Information Behaviour at Shanghai International Studies University (No. 202402280009). It was supported by the Ministry of Education Foundation for Humanities and Social Sciences of China (No. 23XJC740007) and the China Scholarship Council (No. 2406900146). We are grateful to our students who made this work possible, to the research group at the Shanghai International Studies University and the AFO research group at the University of Agder, and to the editors and the reviewers for their insightful suggestions.

References

Albir, P. Group. A. H., Galán-Mañas, A., Kuznik, A., Olalla-Soler, C., Rodríguez-Inés, P., & Romero, L. (2020). Translation competence acquisition. Design and results of the PACTE group's experimental research. *Interpreter and Translator Trainer*, *14*(2), 95-233. https://doi.org/10.1080/1750399X.2020.1732601

https://doi.org/10.26034/cm.jostrans.2026.8284

Association of Chinese Graduate Education. (2024). *The introduction to graduate education disciplines and specialisations and their basic degree requirements.* https://www.acge.org.cn/encyclopediaFront/enterEncyclopediaIndex

Baer, B. J., & Bystrova-McIntyre, T. (2009). Assessing cohesion: Developing assessment tools based on comparable corpora. In C. V. Angelelli, & H. E. Jacobson (Eds.), *Testing and Assessment in Translation and Interpreting Studies: A Call for Dialogue between Research and Practice* (pp. 159-184). John Benjamins. https://doi.org/10.1075/ata.xiv

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1-48. https://doi.org/10.18637/jss.v067.i01

Brooks, M. E., Kristensen, K., Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Mächler, M., & Bolker, B. (2017). glmmTMB balances speed and flexibility among packages for Zero-inflated Generalized Linear Mixed Modeling. *The R Journal*, *9*(2), 378-400.

Bundgaard, K., & Christensen, T. (2019). Is the concordance feature the new black? A workplace study of translators' interaction with translation resources while post- editing TM and MT matches. *The Journal of Specialised Translation*, *31*, 14-37. https://doi.org/10.26034/cm.jostrans.2019.175

Bundgaard, K., Christensen, T. P., & Schjoldager, A. (2016). Translator-computer interaction in action — An observational process study of computer-aided translation. *The Journal of Specialised Translation*, *25*, 106-130. https://doi.org/10.26034/cm.jostrans.2016.302

Carmo, F., & Koponen, M. (2024). Revisers and post-editors: The guardians of quality. In G. Massey, M. Ehrensberger-Dow & E. Angelone (Eds.), *Handbook of the Language Industry - Contexts, Resources and Profiles* (pp. 203-224). De Gruyter Mouton. https://doi.org/10.1515/9783110716047-010

Cheung, K. K. C., & Tai, K. W. H. (2023). The use of intercoder reliability in qualitative interview data analysis in science education. *Research in Science & Technological Education*, *41*(3), 1155-1175. https://doi.org/10.1080/02635143.2021.1993179

China's Language and Terminology Committee. (2021). *Translation services—Post-editing of machine translation output—Requirements* (GB/T 40036—2021/ISO 18587: 2017). https://std.samr.gov.cn/gb/search/gbDetailed?id=C1A814733B1E7A48E05397BE0A0A1C8D

China's Language and Terminology Committee. (2022). *Translation services—Part 1, Requirements for translation services* (GB/T 19363.1—2022/ISO 17100:2025). https://std.samr.gov.cn/gb/search/gbDetailed?id=F159DFC2A8DD47EFE05397BE0A0AF334

Christensen, R. (2025). ordinal: Regression models for ordinal data. R package version 2025.12-29, https://CRAN.R-project.org/package=ordinal.

Chung, E. S. (2020). The effect of L2 proficiency on post-editing machine-translated texts. *The Journal of AsiaTEFL*, *17*(1), 182-193. https://doi.org/10.18823/asiatefl.2020.17.1.11.182

Cui, Y., Zhu, J., Yang, L., Fang, X., Chen, X., Wang, Y., & Yang, E. (2022). CTAP for Chinese: A linguistic complexity feature automatic calculation platform. *Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022)*, *France*, 5525-5538. https://aclanthology.org/2022.lrec-1.592/

Daems, J., Macken, L. (2020). Post-editing human translations and revising machine translations: Impact on efficiency and quality. In M. Koponen, B. Mossop, Robert, S. I., & G. Scocchera. (Eds.), *Translation Revision and Post-editing: Industry Practices and Cognitive Processes* (pp. 50-70). Routledge. https://doi.org/10.4324/9781003096962

Daems, J., Vandepitte, S., Hartsuiker, R., & Macken, L. (2017). Translation methods and experience: a comparative analysis of human translation and post-editing with students and professional translators. *Meta*, *62*(2), 245-270. https://doi.org/10.7202/1041023ar

https://doi.org/10.26034/cm.jostrans.2026.8284

Dai, G., & Liu, S. (2024). Towards predicting post-editing effort with source text readability: An investigation for English-Chinese machine translation. *The Journal of Specialised Translation*, *41*, 206-229. https://doi.org/10.26034/cm.jostrans.2024.4723

Dong, D., & Lan, Y. (2010). Textual competence and the use of cohesion devices in translating into a second language. *Interpreter and Translator Trainer*, *4*(1), 47-88. https://doi.org/10.1080/1750399X.2010.10798797

English Language Teaching Advisory Board under the Ministry of Education. (2020). *Teaching guides for foreign language and literature majors in general, colleagues and universities*. Foreign Language Teaching and Research Press.

Feng, L., & Yan, M. (2020). *Research on China's Standards of English Language Ability Translation Competence Scales*. Higher Education Press.

Frankenberg-Garcia, A. (2022). Can a corpus-driven lexical analysis of human and machine translation unveil discourse features that set them apart? *Target*, *34*(2), 278-308. https://doi.org/10.1075/target.20065.fra

Granger, S., & Lefer, M. (2021). *Translation-oriented annotation system manual Version 2.0*. Centre for English Corpus Linguistics. https://cdn.uclouvain.be/groups/cms-editors-cecl/cecl-papers/TAS-2.0_annotation_manual_2021-10-26.pdf

Guerberof, A. (2014). The role of professional experience in post-editing from a quality and productivity perspective. In S. O'Brien, L. W. Balling, M. Carl, M. Simard & L. Specia (Eds.), *Post-editing of Machine Translation: Processes and Applications* (pp. 51-76). Cambridge Scholars Publishing.

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. Longman Group.

International Organisation for Standardisation. (2024). *Translation services—Evaluation of translation output—General guidance* (ISO Standard No. 5060: 2024). https://www.iso.org/standard/80701.html

Jia, Y. (2024). Integrating translation project management platforms with generative AI technologies: An investigation into the translation process through human-machine interaction. *Foreign Language Teaching and Research*, *6*, 937-949.

Jia, Y., & Sun, S. (2023). Man or machine? Comparing the difficulty of human translation versus neural machine translation post-editing. *Perspectives*, *31*(5), 950-968. https://doi.org/10.1080/0907676X.2022.2129028

Jia, Y., & Zheng, B. (2022). The interaction effect between source text complexity and machine translation quality on the task difficulty of NMT post-editing from English to Chinese: A multi-method study. *Across Languages and Cultures*, *23*(1), 36-55. https://doi.org/10.1556/084.2022.00120

Jia, Y., Carl, M., & Wang, X. (2019). How does the post-editing of neural machine translation compare with from-scratch translation? A product and process study. *The Journal of Specialised Translation*, *31*, 60-86. https://doi.org/10.26034/cm.jostrans.2019.177

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

Károly, K. (2017). *Aspects of cohesion and coherence in translation*. John Benjamins. Kuckartz, U., & Rädiker, S. (2019). *Analysing qualitative data with MAXQDA*. Springer.

Läubli, S., & Green, S. (2019). Translation technology research and human–computer interaction (HCI). In M. O'Hagan (Ed.), *The Routledge Handbook of Translation and Technology* (pp. 370-383). Routledge. https://doi.org/10.4324/9781315311258

https://doi.org/10.26034/cm.jostrans.2026.8284

Läubli, S., Simianer, P., Wuebker, J., Kovacs, G., Sennrich, R., & Green, S. (2022). The impact of text presentation on translator performance. *Target*, *34*(2), 309-342. https://doi.org/10.1075/target.20006.lau

Lei, L. & Zhang, T. (2025-Sep). AlphaReadabilityChinese 2.0. https://github.com/corpustalk/AlphaReadabilityChinese2.0/releases

Lei, L., Wei, Y., & Liu, K. (2024). AlphaReadabilityChinese: A tool for the measurement of readability in Chinese texts and its applications. *Foreign Languages and Their Teaching*, *46*(1), 83-93.

Łoboda, K., & Mastela, O. (2023). Machine translation and culture-bound texts in translator education: A pilot study. *Interpreter and Translator Trainer*, *17*(3), 503-525. https://doi.org/10.1080/1750399X.2023.2238328

Lyu, X., & Feng, L. (2024). The application of the *China's Standards of English Language Ability* translation scales empowered by AI in teaching assessment. *Foreign Language World*, (6), 29-36.

Mann, W., & Thompson, S. (1987). *Rhetorical Structure Theory: A theory of text organization*. Information Science Institute.

Meshkati, N. (1988). Toward development of a cohesive model of workload. *Advances in Psychology*, *52*, 305-314.

Moorkens, J. (2025). The machine translator's visibility: A postphenomenological analysis of machine translation. *Translation Spaces,* https://doi.org/10.1075/ts.23030.moo

Moorkens, J., Toral, A., Castilho, S., & Way, A. (2018). Translators' perceptions of literary post-editing using statistical and neural machine translation. *Translation Spaces*, *7*(2), 240-262. https://doi.org/10.1075/ts.18014.moo

Mu, L., Liang, W., & Liu, X. (2024). China's Standards for Translator and Interpreter Competence Assessment and China's Standards of English Language Ability Translation and Interpreting Competence Scales. *Foreign Languages in China*, *21*(3), 87-97.

Nitzke, J., & Gros, A. K. (2020). Preferential changes in revision and post-editing. In M. Koponen, B. Mossop, I. S. Robert, & G. Scocchera (Eds.), *Translation Revision and Post-editing: Industry Practices and Cognitive Processes* (pp. 21–34). Routledge. https://doi.org/10.4324/9781003096962

O'Brien, S. (2012). Translation as human-computer interaction. *Translation Spaces*, *1*, 101-122. https://doi.org/10.1075/ts.1.05obr

O'Brien, S., Ehrensberger-Dow, M., Hasler, M., & Connolly, M. (2017). Irritating CAT tool features that matter to translators. *HERMES - Journal of Language and Communication in Business*, *56*, 145-162. https://doi.org/10.7146/hjlcb.v0i56.97229

Pan, J., Wong, B. T. M., & Wang, H. (2022). Navigating learner data in translator and interpreter training: Insights from the Chinese/English translation and interpreting learner corpus (CETILC). *Babel*, *68*(2), 236-266. https://doi.org/10.1075/babel.00260.pan

Peng, Y., Hu, R., & Wu, J. (2023). Automatic analysis and application of cohesion features of Chinese texts. *Yuyan Wenzi Yingyong* (*Applied Linguistics)*, *1*, 114-129.

Pietrzak, P., & Kornacki, M. (2021). *Using CAT tools in freelance translation: Insights from a case study*. Routledge. https://doi.org/10.4324/9781003125761

Qian, J., Xiao, W., Li, Y., & Xiang, X. (2022). Impact of neural machine translation error types on translators' allocation of attentional resources: Evidence from eye-movement data. *Foreign Language Teaching and Research*, *5*, 750-761.

Richards, J. C., & Schmidt, R. W. (2011). *Longman dictionary of language teaching and applied linguistics* (4th ed.). Routledge. https://doi.org/10.4324/9781315833835

https://doi.org/10.26034/cm.jostrans.2026.8284

Sánchez-Gijón, P., Moorkens, J.,& Way, A. (2019). Post-editing neural machine translation versus translation memory segments. *Machine Translation*, *33*, 31-59. https://doi.org/10.1007/s10590-019-09232-x

Schaeffer, M. J. (2022). The impact of translation competence on error recognition of neural MT. *Proceedings of the 15th biennial conference of the Association for Machine Translation in the Americas (AMTA 2022) Workshop 1: Empirical Translation Process Research*, *USA*, 41-48. Association for Machine Translation in the Americas. https://aclanthology.org/2022.amta-wetpr.5

Schaeffer, M., Huepe, D., Hansen-Schirra, S., Hofmann, S., Muñoz, E., Kogan, B., Herrera, E., Ibáñez, A., & García, A. M. (2020). The Translation and Interpreting Competence Questionnaire: An online tool for research on translators and interpreters. *Perspectives*, *28*(1), 90-108. https://doi.org/10.1080/0907676X.2019.1629468

Silva, B., Buchicchio, M., Stigt, D. V., Stewart, C., & Moniz, H. (2024). Data-driven Asian adapted MQM typology and automation in translation quality workflows. *The Journal of Specialised Translation*, *41*, 98-126. https://doi.org/10.26034/cm.jostrans.2024.4713

State Language Commission. (2024). *China's Standards of English Ability (2024)*. Shanghai Foreign Language Education Press.

Sun, S., Wang, H., & Jia, Y. (2025). Direction matters: Comparing post-editing and human translation effort and quality. *PLoS One*, *20*(7), e0328511. https://doi.org/10.1371/journal.pone.0328511

Swar, O., & Mohsen, M. (2023). Students' cognitive processes in L1 and L2 translation: Evidence from a keystroke logging program. *Interactive Learning Environments*, *31*(10), 6696-6711. https://doi.org/10.1080/10494820.2022.2043386

Tezcan, A., Daems, J., & Macken, L. (2019). When a 'sport' is a person and other issues for NMT of novels. *Proceedings of the Qualities of Literary Machine Translation*, *Ireland*, 40-49 . https://aclanthology.org/W19-7306/

Venables, B., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). Springer.

Vieira, L. N., Zelenka, N., Youdale, R., Zhang, X., & Carl, M. (2023). Translating science fiction in a CAT tool: Machine translation and segmentation settings. *Translation & Interpreting*, *15*(1), 216-235. https://doi.org/10.12807/ti.115201.2023.a11

Wang, X., Li, X., & Chen, G. (2024). Comparing translation revision and machine translation post-editing: Evidence from keylogging, retrospection and questionnaire. *Foreign Language Learning Theory and Practice*, *5*, 88-97.

Wang, X., Wang, L., & Zheng, B. (2022). Impact of translation directions on information processing and translation quality: A triangulated study using eye-tracking and screen-recording. *Foreign Language Teaching and Research*, *1*, 128-139.

Yao, Y., Han, T., & Li, D. (2025). Measuring translation trainees' effort in AI-assisted post-editing: A multi-method approach. *Interpreter and Translator Trainer*, *19*(3-4), 357-378. https://doi.org/10.1080/1750399X.2025.2535239

Yuan, R. (2022). *The difficulty of teaching materials for interpreter trainees: A text-structured approach*. Shanghai Jiao Tong University Press.

Zhang, Q., Osborne, C., & Moorkens, J. (2024). MT error detection and correction by Chinese language learners. *Translation and Interpreting Studies*, *19*(2), 277-301. https://doi.org/10.1075/tis.22092.zha

https://doi.org/10.26034/cm.jostrans.2026.8284

Data availability statement

List of indicators and results from CTAP/ARC 2.0, and the R Markdown for the model analysis can be found under https://osf.io/hfybc/overview?view_only=23636637768a484c802214fbf0a8ce11

Endnotes

1. The Lan-BridgeMT has been trained on billions of quality parallel texts over 208 languages from the company's 26 years of professional translation services across over 50 domains, such as culture, engineering, and transportation. It outperformed other competitors (e.g., GPT4-5shot) in the document-level shared translation tasks from Chinese to English at the 2023 Conference on Machine Translation.
2. To ensure NMT suggestions were the same for M3 and M4, we checked twice before every participant worked on their tasks.
3. CATTI Level 2 is higher than Level 3.
4. There was one group with one translator and two bilingual revisers.
5. The score thresholds were determined by the value exceeding ±1 *SD* from the *M* value.
6. For the other three themes below this threshold, we have consulted relevant literature and have been developing a more theory-driven code frame.
7. The full explanation of the code system is beyond the scope of this study, which will be reported in another paper in preparation.
8. We appreciate one of the anonymous reviewers' suggestions on the methodology.

https://doi.org/10.26034/cm.jostrans.2026.8284