Article

Comparing Generative AI and Advanced Learners in Translation Tasks Using the House Model: Similarities and Pedagogical Implications

Chenqing Song*
Qifei Kao
Binghamton University, State University of New York, USA

Abstract

This research examines the performance of Generative AI models (Gen AIs) compared to advanced language learners of Chinese in prompted translation tasks. We utilize discourse analysis (Bender & Lascarides, 2019; Cruse, 2011; Goddard, 1998; Kroeger, 2019), especially social-cultural analysis, through the functionalist model of translation quality assessment framework proposed by House (1997, 2014). This model assesses the functional equivalence of the translations through linguistic-textual analysis across four dimensions, i.e., 1) Genre and three classic Hallidayan register categories, 2) Field, 3) Mode, and 4) Tenor. It is often observed that Gen AIs excel at generating error-free sentences due to their extensive linguistic databases, while advanced learners produce few errors by using alternative expressions for gaps of their linguistic knowledge, a communicative strategy, i.e., circumlocution or avoidance (Váradi, 1980; Dörnyei & Scott, 1997; Kasper & Kellerman, 1997). Therefore, the output of writing tasks presents difficulties to close comparison due to its variability. Unlike writing, translation requires precise message conveyance without paraphrasing.

In our study, three advanced Chinese learners and five Gen AI models completed translation of eight texts (each with a word count of 400-500) from English to Chinese. When evaluated with this model, Gen AIs produce awkward or inappropriate translations, failing to capture nuanced tones, styles, and contextual suitability. Similarly, advanced learners occasionally choose expressions that, while seemingly correct, are contextually inappropriate, highlighting the need for refinement in their language proficiency. To leverage these similarities in functional mismatches, we propose a collaborative revision process (Min, 2006; Hu & Lam, 2010; Henshaw, 2023) in advanced language teaching. Learners compare their translations with those generated by Gen AIs and revise them based on the teachers' feedback on discourse and functional issues to improve their linguistic competence. This research presents instructional design ideas, classroom activities, and student assignments for implementing this collaborative revision process effectively.

Keywords

*Corresponding author. Email: csong@binghamton.edu

1 Introduction

1.1 Gen AIs and language learning

The emergence of Generative AIs has significantly transformed the landscape of language education, presenting both challenges and opportunities for language learners and teachers. On the positive side, Gen AIs can provide language learners with immediate feedback, enabling them to correct mistakes in real time. They also allow educators to create immersive and interactive learning environments, where learners can practice language skills through simulated conversations and contextualized scenarios. Moreover, teachers can leverage Gen AIs' adaptability to a learner's proficiency level, providing customized instructional materials that appropriately challenge and support them.

While scholars acknowledge the benefits of incorporating Gen AIs in language instruction (Baidoo-Anu & Ansah, 2023; Jiang et al., 2023; Kohnke et al., 2023; Li et al., 2023; Pérez-Núñez, 2023; Simonsen & Bédib, 2023; Tan et al., 2023), concerns persist regarding potential adverse effects such as overreliance on these tools and decreased human interaction (Zunaidah et al., 2023). Two studies on college students' use of Gen AI in college academic writing (Kim et al., 2024; Mennella & Quadros-Mennella, 2024) have revealed that learners tend to use Gen AIs for output tasks, including writing and translation, raising concerns among scholars and educators about overreliance, misuse, and cheating. Furthermore, the quality and accuracy of AI-generated content can vary, sometimes leading to the reinforcement of incorrect usage or misunderstandings.

To further complicate the picture, there is growing concern about the value of learning a foreign language due to the rapid advancement of Gen AIs' language abilities. It is not difficult to find opinions in traditional media as well as on social media outlets, such as "AI could make it less necessary to learn foreign languages" in the Economist (Johnson, 2023) or AI is "The End of Foreign-Language Education" in the Atlantic (Bosker, 2024). As language teachers, we intuitively resist accepting such predictions, but we must evaluate them with scientific investigations and empirical research.

In recent discussions in the field of Chinese language teaching and learning, scholarly research has explored the advantages of Gen AIs, particularly ChatGPT, in supporting Chinese language learners. Li et al. (2024) demonstrated ChatGPT's efficacy in generating materials across various genres and difficulty levels, creating teaching tasks, and simulating assessments. Meng (2024) assessed ChatGPT's capabilities in teaching vocabulary and concluded that it has yet to match native speakers – especially language teaching professionals – in their intuition and analytical understanding of vocabulary. Song et al. (2024) analyzed the feasibility of generating reading materials for young learners using ChatGPT and Midjourney.

Among this recent scholarship, studies focusing on teaching writing are directly relevant to the research questions posed in the current study. Based on feedback from language instructors, Li et al. (2023) advocated for the use of ChatGPT in teaching academic writing, leveraging its role as a peer in "dialogues" with students. Gao (2025) investigated the similarities and differences between feedback on writing provided by human instructors and by ChatGPT. Gao suggested that ChatGPT is more helpful to advanced learners in writing tasks due to its strength in the stylistic, logical, and discoursal features of the Chinese language. In this paper, we challenge Gao's conclusion through a closer analysis of translation outputs produced by both advanced learners and generative AI tools. Based on this comparison, we argue that the role of Gen AIs as a peer to language learners in output tasks is built not on its superiority, but on the similarities, it shares with learners in terms of stylistic, logical, and discoursal capacities.

1.2 Assessing translation

The current research on Gen AIs' performance in translation tasks primarily originates from the field of Natural Language Processing (NLP), often conducted by research teams that create Gen AI models. Such

studies (Castilho et al., 2023; Hendy et al., 2023; Jiao et al., 2023) commonly use automatic metrics, such as BLEU, chrF, and TER, for evaluation, with or without additional bilingual human-based assessments. Hendy et al. (2023) concluded that "GPT systems can produce highly fluent and competitive translation outputs in the zero-shot setting, especially for high-resource language translations." Similarly, Jiao et al. (2023) compared GPT-4's translation ability with Google Translate, GPT-3.5, and DeepL, finding that "ChatGPT performs competitively with commercial translation products (e.g., Google Translate) on high-resource languages but lags significantly behind on low-resource or distant languages." However, the authors also noted that GPT-4 significantly improves translation performance.

While these studies conducted by NLP specialists closely follow the development in Gen AIs, they also suffer from limitations due to the NLP approach. First, the current state of neural-network-based automatic metrics mostly relies on lexical matching, with the evaluation unit typically being sentences. Hendy et al. (2023) acknowledged this gap and identified that developing metrics to measure contextual correctness is a future direction. Secondly, the human-based evaluations in these studies often involve bilingual individuals without formal training in translation, making them less aware of nuanced linguistic differences and complex linguistic functions. In fact, Jiang and Zhang's (2024) comparative study revealed that when evaluating ChatGPT's translation performance, the automatic metrics and human evaluations diverge, particularly regarding semantic and pragmatic fidelity.

On the other hand, the assessment of human performance in translation tasks follows its own tradition. Moving beyond ad hoc comments and intuitive judgments about the quality of translations, the field of translation studies has developed systematic tools for Translation Quality Assessment (TQA) rooted in various theoretical models of language, communication, and translation to evaluate the quality of translations produced by human translators. Current TQA standards and models are widely used in academia, as exemplified by Williams (2004), House (1997, 2014), and Bittner (2020), as well as those employed in the translation profession or industry, such as Drugan (2013). Professional associations and government certification exams, such as those from the American Translators Association and the China Accreditation Test for Translators and Interpreters, follow their own models, influenced by both scholarly research and industry practices. These TQA models and standards are all human-based and require well-trained evaluators to perform the assessments.

Although these standards and models were originally designed to evaluate human-produced translations, some scholars have also applied human-based TQA to Gen AI models since the debut of ChatGPT, using frameworks such as the Multidimensional Quality Metrics (MQM) error typology. For instance, Pu (2023) assessed the quality of ChatGPT's Chinese-to-English translations from the perspective of translator behavior criticism. Aghai (2024) evaluated literary translations from Persian to English generated by ChatGPT and Google Translate, employing Sofyan and Tarigan's (2019) functional holistic model, and contended that neither system delivers high-quality literary translations capable of effectively convey cultural nuances and idiomatic expressions. Nagi, Alzain, and E. Naji (2023) conducted an error taxonomy analysis on English-to-Arabic translations based on the MQM.

In summary, existing TQA studies of Gen AI translation outputs either compare them to other machine translation models, mostly Neural Machine Translation (NMT) systems such as Google Translate, or to professional translators. However, it remains unknown how Gen AIs perform compared to advanced human learners who are not trained as translators.

1.3 Research questions

Against this backdrop, this paper explores the output of Gen AIs and advanced Chinese learners in translation tasks. Using a comparative analysis grounded in the House model for translations from English to Chinese, we aim to answer the following research questions:

1. How do Gen AIs compare to advanced learners with proficiency levels equal to or higher than HSK Level 6 in translation output tasks?

2. How can teachers leverage the similarities between Gen AIs and advanced learners to facilitate language learning?

Translation and language proficiency are related but distinct abilities, and learning a language involves much more than performing translation tasks. In this study, we selected translation as the evaluative task instead of writing or other language skills because translation demands conveying messages precisely from one language to another without the flexibility of circumlocution, paraphrasing or rephrasing. This also allows a side-by-side comparison between human translators and Gen AIs.

More research is needed to reveal the differences and similarities between Gen AIs and language learners across various aspects of language proficiency. Furthermore, even if Gen AIs outperform human learners in certain areas, the intrinsic value and benefits of learning another language remains when learners' motivation goes beyond mere practicality (Fox et al., 2019). In this context, the comparison between Gen AIs and advanced learners in this study focuses less on determining which is superior and more on identifying potential similarities, which provides a foundation for answering the third research question on examining the pedagogical values of Gen AIs in helping language learners. This study, therefore, compares some of the most popular Gen AIs with advanced learners of Chinese to investigate how these tools can be effectively leveraged in language teaching and learning.

2 Research Method

2.1 The House Model

In this study, among the human-based models, we chose to employ the House model (House, 1997, 2014), one of the most well-known and widely applied TQA models. The House model was constructed based on Hallidayan functional linguistic theory concerning language and communication and is connected to discourse analysis (Bender & Lascarides, 2019; Cruse, 2011; Goddard, 1998; Kroeger, 2019). It aims to evaluate a translation as a functional equivalent to the original, measured by matches or mismatches in particular features of the context situation or "situational dimensions," including Field, Mode, Tenor, and Genre. Linguistic means to achieve matches along these dimensions include syntactic, lexical, and textual (contextual) elements, including sentence connections, thematic movements, and iconic linkage. The House model can be illustrated as follows:

Figure 1.
A Scheme for Analyzing and Comparing Original and Translations Texts (House, 2014, p. 65)
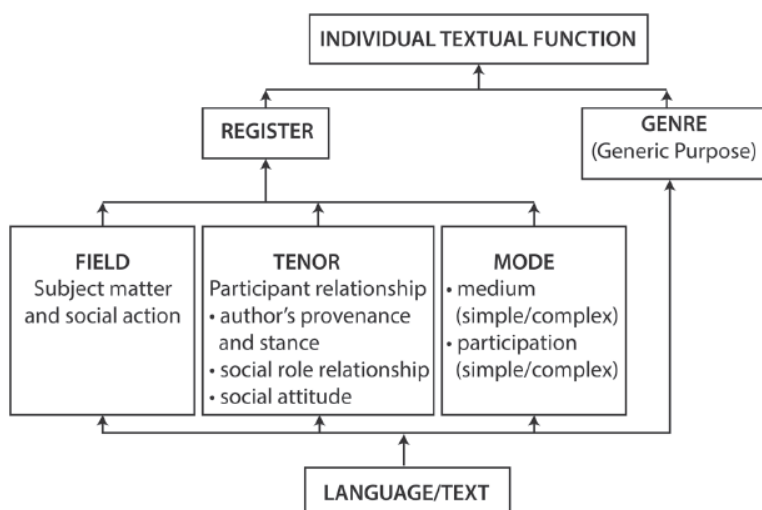
Definitions of the key terms in this scheme are given below (House, 2014, p. 65):
- Field: topic and content of the text or its subject matter.
- Tenor: nature of participants, the addresser and the addresses, their relationship.
  - Author's provenance and stance: text producer's temporal, geographical and social provenance, their intellectual, emotional or affective stance, personal viewpoint
  - Social role relationship: relationship in terms of social power and social distance
  - Social attitude: different styles (formal, consultative, and informal)
- Mode: written vs. oral text and the participation between reader and writer.
  - Medium: "simple medium" means a text is either written or spoken; "complex medium" crosscuts written and spoken, e.g. written to be spoken, or written to be read as if heard.
  - Participation: simple participation means it is a monologue with no addressee participation; complex means various addressee-involving linguistic mechanisms.

Besides functional mismatches, the House model admits that translations may also contain more basic or overt errors, which can be categorized into two types. The first type includes basic denotational mismatches between the source text and the translation, such as omissions, additions, incorrect substitutions, etc. The second type involves errors that breach the target language system. While the House model's strength is not in detecting basic errors (making it potentially less useful for evaluating lower-proficiency translators), this characteristic aligns perfectly with our research needs. Firstly, since we expect both advanced learners of Chinese and Gen AIs to display higher levels of Chinese proficiency with fewer basic errors, the model's focus on functional and discoursal aspects rather than basic linguistic errors is appropriate for our study population. Secondly, the House model connects functional mismatches with linguistic means—syntactic, lexical, or discoursal—which are also units of teaching familiar to language teachers. Thus, translation quality assessment using the House model can potentially offer valuable insights for teaching and learning in the language classroom. Thirdly, the House model has been widely applied in Chinese-English translations done by high-level amateur or professional translators, and its effectiveness has been well-documented (He et al., 2024; Hu, 2018). Finally, the House model is not a quantitative system that assigns a score to translations, nor does it prescribe the "best translation." Its focus on functional matches and mismatches allows us to identify similarities and differences between Gen AIs and Chinese learners. Since our primary concern is not which is better but rather the potential of Gen AIs as tools for teaching and learning activities, we believe the House model offers a more suitable framework than purely quantitative systems.

Bittner's (2020) model is also highly systematic, but it focuses on tracing the decision-making process underlying translation practice, taking into account various factors surrounding a particular translation to inform the most appropriate translation strategy. We believe that tracking the decision-making process goes beyond the scope of the current study and would unnecessarily complicate the analyses. Additionally, we did not select automatic metrics because the human participants in this study could not produce the large volume of output data required for such metrics.

2.2 Selection of the Source Texts

Another advantage of adopting the House model is that its organization guided the researchers in designing the research method and procedure. First, the model, with its taxonomy of the four situational dimensions that make up the overall function of any text, informs the selection of source texts to be translated. Rather than randomly choosing a text, we selected eight texts, each with a word count of 400-500, that differ significantly in three of the situational dimensions. All eight selected texts share one common dimension—namely, the Field, or the broad subject on food—which facilitates the Chinese

learners' completion of the translation tasks more quickly, assuming the texts share some vocabulary. The differences in the other three dimensions, i.e., Tenor, Mode, and Genre, allow researchers to study the effects, if any, of these situational dimensions. The basic information about the eight texts selected for translation is provided below. Four of the eight texts are news articles, and four are excerpts from modern novels. Four of them are simple written texts, while the other four contain oral features (written to be read). In four of the texts, the author-reader or participant is of equal stance, while in the other four, the author or one participant holds a higher or authoritative stance over the reader(s) or other participant(s). No published Chinese translations existed for any of the eight texts before this research. Besides the situational dimensions, we also selected texts that are all recent compositions intended for mainstream public readers[1]. The readability and level of difficulty of the English texts were evaluated using the Flesch-Kincaid Grade Level formula, ensuring that no text is excessively difficult to comprehend. All texts fall within the 8th to 11th grade range on the Flesch-Kincaid scale.

Table 1

The Situational Dimensions of the Eight Selected Texts

	Source	Field	Genre	Mode	Tenor
Text 1	From *The best thing I ate: Smoked Tuna Belly Cake from Fyr and Salt* by Michelina Da Fonte (Da Fonte, 2003)	A food review	News	Written	Equal
Text 2	From *Turns out there's a scientific reason that we always have room for dessert* by Rosemary Slade (Slade, 2023)	A study of human eating behaviors	News	Written	Authoritative
Text 3	From *In New York, a 'cake renaissance' has customers lining up for slices* by Joyce Koh (Koh, 2023)	An exploration of New York City's cake bakeries	News	Spoken	Equal
Text 4	From *Cake in the office should be viewed like passive smoking, obesity expert warns* by Rachel Sylvester, Chris Smyth, and Kat Lay (Sylvester et al., 2023)	The health impacts of workplace sugar consumption.	News	Spoken	Authoritative
Text 5	From *Swann's Way*, by Marcel Proust (Proust, 2014).	A literary exploration of memory	Fiction	Written	Equal
Text 6	From *Blood, Bones, and Butter: The Inadvertent Education of a Reluctant Chef* by Gabrielle Hamilton (Hamilton, 2011)	A Chef's culinary memoir	Fiction	Written	Authoritative
Text 7	From *Chocolat* by Joanne Harris (Harris, 1999)	Contemporary fiction with magical realism	Fiction	Spoken	Equal
Text 8	From *The School of Essential Ingredients* by Erica Bauermeister (Bauermeister, 2009)	A culinary fiction with personal transformation.	Fiction	Spoken	Authoritative

2.3 Research Design and Procedure

The research procedure followed four steps. First, after selecting the source texts, we recruited three advanced Chinese learners with proficiency levels equal to or higher than HSK Level 6[2]. All three had completed a fourth-year university-level Chinese course. Two of the learners (referred to as LC and TL below) were heritage speakers, while one (referred to as KL) was a non-heritage speaker who had spent two years studying abroad in China. None of the learners had prior experience with functional linguistics, discourse analysis, or translation assessment systems.

The learners were instructed to translate the eight texts, using dictionaries if needed, but without assistance from others, machine translation, or Gen AI tools. We made it clear that the translation task was entirely voluntary and unrelated to any class requirements or assignments, and their participation had no impact on their academic standing or grades. We also emphasized that the task was a fun learning adventure and that they would gain insights into common mistakes in AI translation, learn to avoid them, learn to leverage AI effectively and improve their translation skills with professional feedback, which reinforced that they should not rely on AI for this task. Additionally, the translation process was monitored by the instructor to ensure that students only used online resources for information and vocabulary lookup. Due to the large word count of the eight texts, students completed their translations over a two-week period, working in the instructor's office on their personal computers after class.

While the human subjects were translating, we also tasked five widely used and accessible Gen AI (LLM) tools, i.e., ChatGPT-3.5, ChatGPT-4o, Google Gemini, Baidu, and Kimi Moonshot, to translate the same eight texts. These tools were selected for their popularity and widespread availability, with most offering free access (except ChatGPT-4o, which has limited free usage), which is important for pedagogical applications in the language classrooms. We included both English-based (ChatGPT-3.5, ChatGPT-4o, Google Gemini) and Chinese-based (Kimi Moonshot, Baidu) Gen AI platforms to provide a balanced comparison of AI translation performance from different linguistic and cultural contexts. All AI translations were conducted on April 23, 2024, reflecting the state of their capabilities at that point in time. The third step involved analyzing the translations from both the human subjects and the Gen AIs using the House model, focusing on the dimensions of Field, Tenor, Mode, and Genre, as well as the linguistic means each translator employed to realize the textual functions. Both the human and machine subjects were given a simple prompt (Appendix) that included key background information about the Field and Genre. The translations collected from both human and machine subjects were then analyzed by the two authors of this paper and another scholar of foreign language teaching and discourse analysis, all trained professionally and/or academically in Translation Studies. All three human evaluators speak Chinese as their native language and English as their professional language. The process of valuation follows the procedure (House, 2014, pp. 75-95), illustrated below, by the two authors individually and then discussed.

- Step 1: Analyze the English source texts and create a functional statement for their genre, field, tenor and mode.
- Step 2: Thoroughly code the source texts and identify the lexical, grammatical, textual/discoursal means for each of the functional dimensions.
- Step 3: Compare the source texts and the Chinese translation and identify the functional mismatches in each of the functional dimensions in the lexical, grammatical, textual/discoursal categories.

To ensure methodological rigor and minimize subjectivity, we only counted clear cases of functional disparity; borderline or ambiguous instances were excluded from analysis. Each potential mismatch required unanimous agreement from all three evaluators to be included in our findings.

3 Results and Discussion

The analysis of the translations of the eight texts, using the assessment scheme outlined in the House model, revealed that functional mismatches occur across all situational dimensions, both by human translators and Gen AIs. In this sense, advanced learners of Chinese and Gen AIs perform very similarly. In the translation data, we identified all 21 mismatch types from both human subjects and Gen AIs, however, this study does not consider the quantity of mismatches, as this is not central to the House model. Instead, the focus is on the types and quality of mismatches. The following table presents all 21 types of mismatches, each labeled with two letters, and, if applicable, a number when the situational dimension has subcategories. The two letters represent the initials of the situation dimensions (from the rows) and the linguistic means (from the columns). For example, a mismatch in the subject Field between the English source text and the Chinese translation, realized through lexical means, is labeled as FL (Field-Lexical mismatch). We listed the number of mismatches made by both human and Gen AIs in their translations of the eight texts under each mismatch label in the table. We used abbreviations to represent each of them. KL, LC and TL are three human Chinese learners and $C_{3.5}$ stands for ChatGPT-3.5, C_{4o} for ChatGPT-4o, GG for Google Gemini, BE for Baidu Ernie Bot, and KM for Kimi Moonshot. Some lexical mismatches also contribute to syntactic or textual mismatches thus double count exists. Although numbers are important, the small subject size and small text corpus make the quantitative information not as relevant. The important discovery of this study is that similar types of functional mismatches are created by both Gen AIs and advanced language learners of Chinese.

The discussion of each case begins with a functional analysis of the text, followed by a close diagnosis of the mismatch in the particular sentence. We believe the four cases that we chose to present are representative, as they involve Field, Mode, and Tenor and all three types of linguistic means. According to House, "Genre connects texts with the 'macro-context' of the linguistic and cultural community in which the text is embedded." It is interconnected with the other three functional dimensions on a macro level; therefore, we did not present any case with Genre mismatch. In each case discussion, we include the context and the sentence where mismatches occur. While there may be incorrect translations, there is no single best translation. In this study, we do not provide an ideal translation as the goal of peer editing is not to achieve a perfect rendition.

Table 2

Twenty-one types of mismatches

		Lexical means	**Syntactic means**	**Textual means**
Field	Subject field	FL *(Human: KL 5; LC 3; TL 4 Gen AIs: $C_{3.5}$ 2; C_{4o} 1; GG 2; BE 1; KM 2)*	FS *(Human: KL 2; LC 0; TL 1 Gen AIs: $C_{3.5}$ 1; C_{4o} 0; GG 1; BE 0; KM 0)*	FT *(Human: KL 1; LC 0; TL 0 Gen AIs: $C_{3.5}$ 0; C_{4o} 0; GG 0; BE 0; KM 0)*
Tenor	1. Author's provenance and personal stance	TL1 *(Human: KL 7; LC 6; TL 6 Gen AIs: $C_{3.5}$ 7; C_{4o} 7; GG 6; BE 8; KM 7)*	TS1 *(Human: KL 3; LC 1; TL 2 Gen AIs: $C_{3.5}$ 3; C_{4o} 3; GG 3; BE 4; KM 3)*	TT1 *(Human: KL 2; LC 0; TL 1 Gen AIs: $C_{3.5}$ 1; C_{4o} 2; GG 2; BE 1; KM 2)*
	2. Social relationship	TL2 *(Human: KL 12; LC 7; TL 9 Gen AIs: $C_{3.5}$ 11; C_{4o} 12; GG 12; BE 13; KM 13)*	TS2 *(Human: KL 9; LC 7; TL 8 Gen AIs: $C_{3.5}$ 8; C_{4o} 8; GG 8; BE 8; KM 7)*	TT2 *(Human: KL 7; LC 5; TL 5 Gen AIs: $C_{3.5}$ 6; C_{4o} 5; GG 6; BE 6; KM 5)*

	3. Social attitude	TL3 *(Human: KL 5; LC 2; TL 3* *Gen AIs: $C_{3.5}$ 3; C_{4o} 3;* *GG 2; BE 3; KM 3)*	TS3 *(Human: KL 3; LC 2; TL 2* *Gen AIs: $C_{3.5}$ 2; C_{4o} 2;* *GG 1; BE 2; KM 1)*	TT3 *(Human: KL 5; LC 3; TL 4* *Gen AIs: $C_{3.5}$ 1; C_{4o} 1;* *GG 1; BE 1; KM 1)*
Mode	1. Medium	ML1 *(Human: KL 6; LC 2; TL 3* *Gen AIs: $C_{3.5}$ 5; C_{4o} 5;* *GG 4; BE 5; KM 5)*	MS1 *(Human: KL 3; LC 2; TL 3* *Gen AIs: $C_{3.5}$ 4; C_{4o} 4;* *GG 3; BE 4; KM 4)*	MT1 *(Human: KL 3; LC 2; TL 2* *Gen AIs: $C_{3.5}$ 1; C_{4o} 1;* *GG 1; BE 1; KM 1)*
	2. Participation	ML2 *(Human: KL 2; LC 2; TL 1* *Gen AIs: $C_{3.5}$ 3; C_{4o} 3;* *GG 3; BE 2; KM 3)*	MS2 *(Human: KL 3; LC 2; TL 3* *Gen AIs: $C_{3.5}$ 2; C_{4o} 2;* *GG 3; BE 2; KM 3)*	MT2 *(Human: KL 3; LC 1; TL 2* *Gen AIs: $C_{3.5}$ 2; C_{4o} 2;* *GG 3; BE 2; KM 2)*
Genre	Generic purpose of the text	GL *(Human: KL 15; LC 11; TL 14* *Gen AIs: $C_{3.5}$ 14; C_{4o} 14;* *GG 13; BE 14; KM 14)*	GS *(Human: KL 13; LC 13; TL 13* *Gen AIs: $C_{3.5}$ 12; C_{4o} 12;* *GG 12; BE 12; KM 12)*	GT *(Human: KL 12; LC 9; TL 11* *Gen AIs: $C_{3.5}$ 10; C_{4o} 11;* *GG 10; BE 12; KM 10)*

Note: KL, LC and TL are three human Chinese learners and C3.5 stands for ChatGPT-3.5, C4o for ChatGPT-4o, GG for Google Gemini, BE for Baidu Ernie Bot, and KM for Kimi Moonshot.

3.1 Case 1: Mismatch Field-Lexical (FL)

Source Text 1. Excerpt from "The best thing I ate: Smoked Tuna Belly Cake from Fyr and Salt" by Michelina Da Fonte:

> That first bite was divine as were all the others to follow until the plate was clean. The tuna belly cake was perfectly prepared with smoky and sweet flavors balanced well by pickled red onions, hot peppers, cucumber and sprigs of arugula. The herbed yogurt gave a bright, creamy, and cooling effect that rounded out the dish.

Field. This excerpt is from a Northforker.com article, a descriptive food review written to engage readers and share a culinary experience.

Tenor. The author's language is unmarked, contemporary, and standard middle-class American English. While the specific background of the author is not detailed in the article, it can be inferred that the author has expertise or a significant interest in food and dining. The author positions themselves as an informed and passionate guide, eager to share a culinary discovery with the readers. The relationship is equal, friendly, and inviting, aiming to create a sense of trust and shared enthusiasm for food exploration. There is no evidence of an educational, pedagogical, or lecturing tone.

Mode. The news article is primarily written for reading but includes some quotes from conversations in an oral yet formal speech style. The excerpt comes from the written-for-reading section of the article. The article is mostly a monologue with minimal dialogic parts, so reader participation is simple.

Genre. The article focuses on a single dish from a particular restaurant and falls under the category of culinary journalism. Its style is characterized by a descriptive and engaging tone meant to convey the sensory experiences of the dish, using evocative language to stimulate the reader's imagination and appetite.

The source text sentence poses difficulties for translators due to a set of adjectives—"bright," "creamy," and "cooling"—that are unique to the subject field of food and cooking. While "bright" is a multisensory description, achieving a functional match in the Field requires translators to select corresponding words in Chinese that are specific to the sense of taste with a positive aesthetic connotation for culinary excellence. Many human translators and Gen AIs rendered it as "明亮," "鲜明," or "色泽丰富," thereby mismatching the sense of taste with the sense of vision. While the Chinese adjective "奶油般的" does stay within the Field, it conflicts with the subject of this sentence: yogurt "酸奶" is a milk product and therefore cannot be described as "milk-like." The English verb "round out" depicts the completeness of the taste, borrowing from the visual perception of spheres, but ChatGPT-4o and Gemini chose to stay within the visual perception by translating it as "圆," which in Chinese, cannot be applied to describe a taste.

Table 3

Translations of the sentence "The herbed yogurt gave a bright, creamy, and cooling effect that rounded out the dish."

Translator	Translation
ChatGPT-3.5	香草酸奶给人清新、丰满和冷却的感觉，完美地衬托了这道菜。
ChatGPT-4o	香草酸奶带来了明亮、奶油般的凉爽效果，使整个菜品更加圆满。
Google Gemini	草本酸奶酱带来明亮、柔滑和清爽的口感，使这道菜更加圆润。
Baidu Ernie Bot	香草酸奶带来了明亮、奶油般的清凉效果，让这道菜更加完美。
Kimi Moonshot	香草酸奶增添了明亮、醇厚、清凉的效果，使整道菜更加完美。
Student KL	加上调味过的酸奶做到了色香味俱全，给这道菜添加了色泽丰富的美感，以及清凉和带有奶油味的口感。
Student LC	再配上香草酸奶清新醇厚的味道，让这道菜成为一道美味的佳肴。
Student TL	酸奶香草酱用一种鲜明，似奶油，以及凉爽的感觉丰富了这道菜。

3.2 Case 2: Mismatch Mode-Syntactic (MS)

Source Text 7. Excerpt from Chocolat by Joanne Harris

> Vianne: I'm glad you approve, Armande. There's nothing quite like the pleasure of indulging in a delicious dessert, especially when shared with good company.
> Armande: Indeed. And speaking of good company, I'm so grateful to have you here, Vianne. You've brought a bit of sweetness into my life that I didn't know was missing.

Field. This excerpt is a dialogue from Joanne Harris's 1999 novel *Chocolat*, a work of contemporary fiction that explores themes of indulgence, community, and transformation. The dialogue between Vianne and Armande highlights the pleasure of enjoying a delicious dessert and the value of good company.

Tenor. The language in this excerpt is contemporary and clear, using standard British English with a warm and inviting tone. Joanne Harris, an English-French writer known for her works that delve into human relationships and sensory experiences, brings her distinctive style to this text, blending warmth, humor, and a touch of mysticism. The interaction between Vianne and Armande reflects a friendly and equal relationship. Their conversation is intimate and respectful, highlighting mutual appreciation and affection, and underscoring themes of community and connection. The dialogue is informal and consultative, with a conversational tone that suggests familiarity and ease.

Mode. Though the novel is primarily written for reading, the excerpt takes the form of spoken dialogue, with characters engaging in direct conversation. This creates a complex medium, incorporating elements typical of spoken language, such as direct address and a conversational tone, in a text designed to be read as if heard. The dialogue between Vianne and Armande has an informal and natural speech style. In terms of participation, the text remains a monologue with no direct reader involvement. The characters speak to each other, and the reader's role is to observe and interpret their interaction. There are no mechanisms that explicitly involve the reader, thereby keeping them as an outside participant who is indirectly engaged through the narrative.

Genre. The excerpt is categorized as contemporary fiction. The style combines narrative and dialogue to explore human relationships and sensory experiences. The language is engaging and evocative, drawing readers into the characters' world and emphasizing themes of community and transformation. The use of dialogue adds intimacy and immediacy, allowing readers to feel the warmth and connection between the characters.

The challenges for translators arise from the syntactic differences between Chinese and English, particularly regarding long relative clauses and their role in the written language. In the source English text, the long and complex relative clause "that I didn't know was missing" is allowed in informal spoken language. In Chinese, relative clauses are formed as a modifier-head structure and are therefore often shorter. Long and complex relative clauses are rarely used in spoken Chinese (Chao, 1968; Wang, 1984 a, b). Most human and Gen AI translators in our study formed a Chinese relative clause and rendered the English sentence as a long ' 的 ' structure, with the nominal head in the sentence-final position, resulting in a sentence unsuitable for the medium. Interestingly, ChatGPT-3.5 was an exception to this problem, but its upgraded version, ChatGPT-4o, exhibited it.

Table 4

Translations of the sentence "You've brought a bit of sweetness into my life that I didn't know was missing."

Translator	Translation
ChatGPT-3.5	你为我的生活带来了一些甜蜜，我之前并不知道缺少了这一点。
ChatGPT-4o	你为我的生活带来了一种我不知道自己缺少的甜蜜。
Google Gemini	你为我的生活带来了一丝我从未意识过的甜美。
Baidu Ernie Bot	你给我的生活带来了我之前不知道缺少的甜蜜。
Kimi Moonshot	你为我的生活带来了一些我未曾意识到缺失的甜蜜。
Student KL	你的出现给我的生命带来了一点甜，而且我自己都没发觉。
Student LC	你为我的生活填补了那块我都没意识到缺失已久的甜蜜。
Student TL	你为我的生活带来了一丝我都不知道缺少的甜。

3.3 Case 3: Mismatch Mode-Textual (MT)

Source Text 8. Excerpt from The School of Essential Ingredients by Erica Bauermeister

"IN MY OPINION, a cake is a lot like a marriage," Lillian began, as she brought eggs, milk, and butter from the refrigerator and put them on the counter. "Admittedly, I don't have a lot of experience," she remarked, holding up her ringless left hand with a wry expression on her face, "but I've often thought that it would be a great idea for couples to make their own wedding cakes, as part of the preparation for their life together. Maybe not so many couples would

end up getting married" - Lillian smiled - "but I think those that did might approach it a bit differently."

Field. The excerpt is from *The School of Essential Ingredients* by Erica Bauermeister, a novel that delves into the transformative power of food and cooking. The passage features a reflective and metaphorical exploration of cake-making and marriage, delivered by Lillian, a character who uses culinary experiences to convey deeper life lessons. The text combines personal insight with culinary philosophy, illustrating how cooking can serve as a lens for understanding relationships and personal growth.

Tenor. The author's language is informal, reflective, and conversational, aiming to engage readers through personal insight and culinary metaphor. The author's stance is introspective and empathetic and uses Lillian's metaphor to explore broader themes of relationship dynamics and personal growth. The relationship between Lillian, the speaker, and her audience (the cooking class participants) is one of mentorship and shared experience. Lillian's role as a teacher allows her to offer personal insights and reflections from a position of authority, but in a friendly, accessible manner, fostering a sense of closeness and mutual understanding with her students. The tone is informal and conversational, reflective of a consultative style that encourages engagement and introspection.

Mode. The passage reflects an internal monologue by Lillian, who shares her personal thoughts and metaphorical insights about cake-making and marriage. Although the text is written, it is crafted to feel like spoken dialogue, with Lillian directly addressing her thoughts to the readers in a reflective, informal manner. The medium of the excerpt is complex, as it is written to be read by the readers as if heard, making the reader feel as though they are the audience of a live dialogue. Participation is simple in this excerpt, as it consists of a monologue where Lillian shares her reflections and insights. While readers are invited to ponder Lillian's thoughts and metaphors, there is no immediate interaction or feedback; the exchange remains unidirectional.

Genre. The excerpt is part of literary fiction that incorporates culinary elements. It features a reflective and metaphorical style, using cooking as a backdrop to delve into themes of personal development and relationships. The passage blends narrative storytelling with philosophical reflections, employing evocative language to connect culinary experiences with broader life insights.

When translating this paragraph into Chinese, translators must handle the complex Mode where a written text, intended to be read silently, is crafted to feel as if it is heard. In the original text, the novelist uses capitalized letters for "IN MY OPINION" to convey a sense of personal emphasis and authority. However, this visual signal must be processed by the readers first as emphatic before the message can be "heard" as part of the conversation recreated by their minds. In Chinese, emphasis in such a complex written-spoken medium is typically conveyed through lexical, syntactic, discoursal, or formatting means (less commonly used) since capitalization is unavailable in the language. In this study, when translating "IN MY OPINION" into Chinese, Gen AI models and human translators all failed to utilize proper Chinese means to match the emphatic tone in the complex Mode. What is intriguing in the data is how student LC translated the expression. She used two Chinese commas to separate the three Chinese characters, of which the last two form a single morphological word. When asked about her creative rendition, she explained that the commas were meant to represent the pauses between the syllables in speech for the emphatic tone. It is clear that she decoded the English text with its complex written-spoken medium and made an effort to recreate the sound to be heard in the Chinese text. In her translation process, the functional mismatch occurred in the coding process, where she tried to recreate the sound effect of the slow speech and pauses in the emphatic tone. The correct match should be matching the emphatic discoursal function, not the sound effect. The schemes below illustrate the functional matching processes.

Figure 2
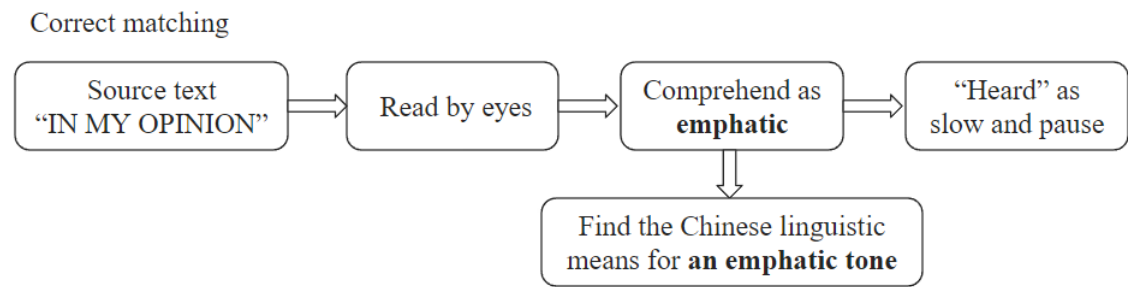
A Scheme for the Correct Matching

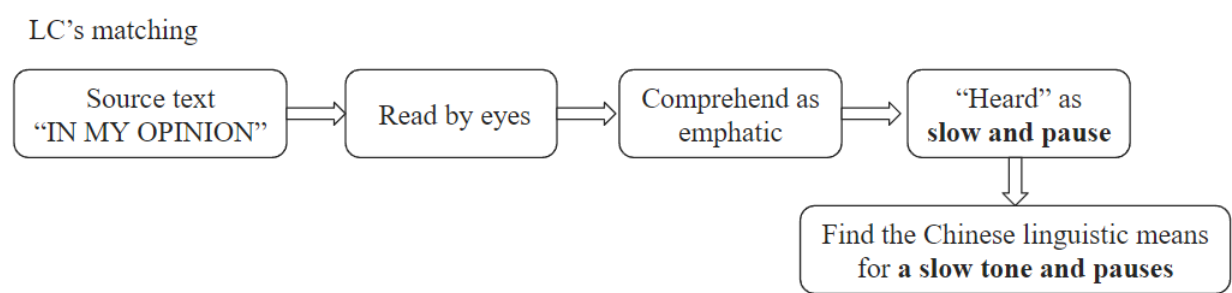Figure 3

A Scheme for LC's Matching

Table 5

Translations of the sentence "IN MY OPINION, a cake is a lot like a marriage," Lillian began, as she brought eggs, milk, and butter from the refrigerator and put them on the counter."

Translator	Translation
ChatGPT-3.5	"在我看来，蛋糕很像一段婚姻，"莉莲说道，她从冰箱里拿出鸡蛋、牛奶和黄油，放在柜台上。
ChatGPT-4o	"在我看来，一个蛋糕很像一段婚姻，"莉莲开始说，一边从冰箱里拿出鸡蛋、牛奶和黄油放在柜台上。
Google Gemini	莉莲一边从冰箱里拿鸡蛋、牛奶和黄油放在柜台上，一边说道："依我看，蛋糕就和婚姻很像。"
Baidu Ernie Bot	"在我看来，蛋糕和婚姻很像，"莉莲边说边从冰箱里拿出鸡蛋、牛奶和黄油，放在厨房的料理台上。
Kimi Moonshot	"在我看来,蛋糕很像婚姻,"莉莲开始说,她从冰箱里拿出鸡蛋、牛奶和黄油,放在柜台上。
Student KL	莉莉安一边把鸡蛋、牛奶和黄油从冰箱拿出来放柜台上，一边说："在我看来，蛋糕就像婚姻一样"
Student LC	"我、觉、得——蛋糕宛如一场婚姻。"莉莉安把从冰箱里拿出来的鸡蛋、牛奶和黄油放到柜台上。
Student TL	"依我看，蛋糕就像婚姻，"莉莉安说到，同时她从冰箱里拿出鸡蛋，牛奶，和黄油放到柜台上。

3.4 Case 4: Mismatch Tenor-Syntax (TS)

Source Text 6. Excerpt from Turns out there's a scientific reason that we always have room for dessert by Rosemary Slade

> If you think back to those truffle French fries you were desperate to taste, the first bite may have been amazing, the second bite was probably just as good, and perhaps even the fifth or sixth. But eventually, that feeling of enjoyment begins to diminish.

Field. This excerpt is from a lifestyle article in a magazine that explores the psychology of eating, specifically focusing on sensory-specific satiety and its impact on our dining experiences. The text is part of a broader discussion about why we continue to crave dessert even after feeling full from a meal. Written in an engaging and relatable style, the article combines scientific insights with vivid descriptions of food to connect with readers and enhance their understanding of the subject.

Tenor. The article's language is casual and accessible, reflecting contemporary middle-class Australian English. The author quotes a researcher on the topic, Dr. Barbara Rolls, whose language is authoritative, rooted in her expertise in human ingestive behavior. Her clear and precise statements, supported by scientific research and specific terms like "sensory-specific satiety," lend credibility to her contributions. Overall, the author uses relatable language and examples to connect with the reader on a personal level, making the content feel more informative than lecturing.

Mode. The excerpt is written for reading as part of a larger text that conveys information in a conversational manner. The formal scientific language is crafted to mimic a spoken tone, enhancing its approachability and engaging the reader. The medium is complex, as the text, though written, is designed to feel as if it could be heard in a public interview. This is achieved through the use of relatable examples, which create a sense of immediacy and direct communication. Participation is complex: while it is mostly a monologue, the use of the pronoun "you" invites the reader into the interaction, encouraging them to mentally engage with the content and relate it to their own experiences.

Genre. The excerpt is a piece of lifestyle writing focused on food psychology, exploring the science behind eating habits, particularly the phenomenon of sensory-specific satiety. The article blends scientific explanation with everyday experiences and is presented in an engaging tone, making the topic accessible and appealing to a wide audience.

The translators faced challenges in matching the Tenor, specifically the author/speaker's stance. Here, in a non-pedagogical but informative tone, the author is giving a polite suggestion, which was appropriately realized in the "if…then" conditional clause, a syntax structure suitable and commonly used in such cases in English. The majority of the human and Gen AI translators in our study matched the syntax structure with its Chinese counterpart, the "如果" conditional clause, resulting in a rigid and mechanical tone that is improper for the informative stance of the author. The proper rendition for the best functional match was given by Google-Gemini, where an imperative sentence form is paired with "一下," which softens the tone but keeps the suggestion intact. Kimi Moonshot also translated it as a subjectless imperative sentence but without "一下," which may be understood as a suggestion or something else. Student LC's translation helps us understand her translation process and choices. LC also added "一下," indicating that she was aware of the sentence's function as a polite suggestion, but she mistakenly believed that the "当" clause (a when-clause) in Chinese also has the function of giving suggestions. Like in Case 3, LC's mismatch problem occurred not in the decoding step but in the coding step into proper Chinese linguistic forms.

Table 6

Translations of the sentence "If you think back to those truffle French fries you were desperate to taste, the first bite may have been amazing, the second bite was probably just as good, and perhaps even the fifth or sixth."

Translator	Translation
ChatGPT-3.5	如果你回想起那些你迫不及待要尝试的松露薯条，第一口也许美味无比，第二口可能同样美味，甚至第五或第六口也是如此。
ChatGPT-4o	如果你回想起那些你曾经迫不及待想尝试的松露薯条，第一口可能非常美味，第二口可能同样好吃，甚至第五口或第六口也一样。
Google Gemini	回想一下当时迫不及待想尝的松露薯条，第一口也许令人惊艳，第二口也同样美味，甚至第五六口都一样好吃。
Baidu Ernie Bot	如果你回想一下那些你迫不及待想尝一尝的松露薯条，第一口可能很惊艳，第二口可能同样美味，甚至第五口或第六口。但最终，那种愉悦感开始逐渐消退。
Kimi Moonshot	回想那些你急切想要品尝的松露薯条，第一口可能非常美妙，第二口可能同样美味，甚至第五口或第六口也是如此。但最终，那种享受的感觉开始逐渐减少。
Student KL	如果你回想那垂涎欲滴、香气扑鼻的炸薯条，你咬下去的第一口一定是回味无穷的。第二口平平无奇，第三口、第四口可能也会一样。
Student LC	当你回想一下你渴望品尝的松露薯条时，你会发现第一口应该很美味。接着第二口也差不多。甚至第五、第六口也都不错。
Student TL	如果你回想到那些你急切渴望想尝到的黑松露薯条，第一口可能色香味浓，第二口应该也差不多，甚至第五第六口依然如此。

4 Pedagogical Implications

Results from the comparative analyses in Section 3 of the paper highlight that while both Gen AI models and advanced learners of Chinese demonstrate high proficiency in the two languages, they face similar types of challenges when trying to maintain situational functional equivalence using proper linguistic means. The results also suggest that the House model of TQA is effective in revealing mismatches that often elude even experienced readers and language teachers. Based on these findings, we offer the following pedagogical suggestions for teachers of advanced Chinese language classes. Although some human subjects and some Gen AIs may outperform others in terms of quantity, it is observed that humans and Gen AIs alike translate English to Chinese with the same types of functional mismatches in various contexts.

4.1 Applying the House Model to teach reading and writing

The House model, although initially developed as a TQA tool in translation studies, has theoretical foundations that are also highly relevant to language teaching. The model is simple, systematic, and conceptually explicit, making it suitable for teachers who have not been trained in functional linguistics. The idea of matching and mismatching linguistic materials and situational functions, which could exist intralingually or interlingually, so it can be adapted for tasks beyond translation, such as reading and writing. The basic design of the House model is to check whether the same functions, realized by the linguistic materials from words to programmatic elements in the source text, are retained in the target texts, realized by the target language linguistics materials. Reading, especially close reading, may be viewed as the first half of a translation process. In reading tasks, a teacher may select texts that pose

challenges due to their special features in any of the functions and guide students when comparing different linguistic means that the author utilizes to deliver such functions. For example, when assigning students to read a newspaper article, teachers may ask them to identify all the words that convey a formal and supportive tone and explain how these words reflect the author's viewpoint and social stance. Teachers can also create targeted comprehension questions or have students complete a function analysis report, as shown below. Please note that the report template provided is comprehensive but might be overwhelming for students new to the House model. In practice, we recommend teachers introduce categories gradually or selectively based on students' familiarity with the model. Teachers might also adapt the activity by asking students to find additional examples within one or more categories. For intermediate to advanced learners, Chinese can be used as the instructional language to enhance immersion.

Table 7

Functional Analysis Report after Reading a Text

Student Name: _____   Date: _____

Title: _____   Text Type/Genre: _____

1. Field: Content and Topic

    a. What is the main topic of the text? Briefly summarize in Chinese.

    Answer: _____

    b. What key words help identify the topic? List three.

    Answer: _____

2. Tenor: Participants and Their Relationship

    a. Author's Provenance and Stance

    Can you infer the author's temporal, geographical, or social background from the text? List one example to support your answer. Highlight as many additional examples as you can.

    Answer: _____

    What is the author's emotional or intellectual stance? List one example to support your answer. Highlight as many additional examples as you can.

    Answer: _____

    b. Social Role Relationship

    What is the relationship between the author and the reader? List one example to support your answer. Highlight as many additional examples as you can.

    Answer: _____

    Does the text indicate power dynamics or social distance? List one example to support your answer. Highlight as many additional examples as you can.

    Answer: _____

c. Social Attitude

Is the text's style formal, consultative, or informal? List one example to support your answer. Highlight as many additional examples as you can.

Answer: _____

3. Mode: Medium and Participation

a. Medium

Is the text written, spoken, or a combination? List one example to support your answer. Highlight as many additional examples as you can.

Answer: _____

b. Participation

Does the text involve direct interaction with the reader, or is it a monologue? List one example to support your answer. Highlight as many additional examples as you can.

Answer: _____

Applying the House model to the writing tasks, to some degree similar to the second half or the output of a translation process, then we can assess the quality of a writing by asking whether the intended functions are realized by the proper linguistic materials. Teachers can assign writing tasks with specific instructions on the intended commutative functions, in terms of the genre, field, mode and tenor. For example, writing assignments may require the students to write a casual conversation between a teacher and a student about their favorite city. Students then can analyze the intended functions in the instruction or even ask Gen AIs to do the analysis for them before they can do so by themselves with more experience. Below is the functional analysis done by ChatGPT-4o, based on the instruction[3].

Genre: The conversation has a casual, conversational nature, fitting the genre of informal dialogue between a teacher and a student.

Tenor: The tone is friendly and respectful, with the teacher showing genuine interest in the student's thoughts, and the student engaging openly in the conversation.

Field: The conversation focuses on the student's favorite city, discussing topics like culture, food, neighborhoods, and the possibility of living there. The field is centered on personal preferences related to travel and cultural exploration.

Communicative Mode: The conversation mimics natural spoken language. The flow is informal, with both parties asking questions and making comments that move the discussion forward in an easy, conversational way.

Then students can start their writing, using the language building blocks to achieve these functions. In the writing assessment stage, teachers may create rubric and grading criteria using the House model. The table below is a sample rubric for grading writing tasks,

Table 8

Rubric developed from House model for writing tasks

Criteria	Excel	Good	Satisfactory need slight revision	Needs significant improvement
	Earn a total score (vocab, grammar, discourse) of 14 or above	Earn a total score of 11 to 13	Earn a total score of 8 to 10	Earn a total score of 1 to 7
Genre: Informal Dialogue Vocab: 1 2 3 4 5 Grammar: 1 2 3 4 5 Discourse: 1 2 3 4 5	Writing fully reflects a natural, informal dialogue. Everyday expressions and turn-taking are used appropriately.	Mostly conversational with minor inconsistencies (e.g., overly formal language).	Some conversational features but feels largely scripted or formal.	Does not resemble an informal dialogue; lacks conversational structure.
Tenor: Friendly Tone Vocab: 1 2 3 4 5 Grammar: 1 2 3 4 5 Discourse: 1 2 3 4 5	Tone is consistently friendly, respectful, and engaging. Teacher's interest and student's openness are evident.	Tone is mostly friendly and respectful, with minor lapses in engagement.	Limited friendliness or respect; moments of awkwardness or detachment.	Tone is inappropriate, unfriendly, or disengaged.
Field: Content Depth Vocab: 1 2 3 4 5 Grammar: 1 2 3 4 5 Discourse: 1 2 3 4 5	Rich content covering culture, food, neighborhoods, and living possibilities with thoughtful details.	Relevant content but lacks depth in one or two areas; some underdevelopment.	Touches on required topics but is superficial or repetitive.	Minimal, irrelevant, or off-topic content.
Communicative Mode Vocab: 1 2 3 4 5 Grammar: 1 2 3 4 5 Discourse: 1 2 3 4 5	Dialogue flows smoothly with logical turns; questions/ comments drive the conversation forward.	Generally flows well, with minor disruptions or slightly unnatural exchanges.	Lacks smooth flow; noticeable awkward or forced transitions.	Disjointed, unnatural, or difficult to follow.

Students can apply the model to evaluate and revise their own writing or analyze a text in Chinese by examining the relationship between the linguistic materials and the situational functions. Similarly, the teacher may create a semi-structured writing task by extracting portions of a news article and having students complete them to mimic the author's tone by using the appropriate linguistic features.

4.2 Leveraging Gen AIs for peer editing in collaborative projects

Peer editing/review has strong theoretical support in the field of language teaching and learning. Studies show that peer editing promotes a learner-centered environment, where students with heightened awareness take on more responsibility and engage in deeper cognitive processing when revising their work based on peer feedback (Min, 2006; Hu & Lam, 2010; Henshaw, 2023). Research consistently shows that peer editing positively influences writing development and encourages collaborative learning (Rollinson, 2005). Despite these benefits for language learning, practical challenges arise when implementing peer editing in the classroom. In small classes – a common scenario in advanced-level Chinese courses – students may have limited opportunities to review a diverse range of writing or translations. Another challenge is that students can often easily identify the author of a piece in such settings, which may lead to peer tension and potential conflicts.

Integrating Gen AIs as collaborators provides solutions to the challenges outlined above. One benefit of employing Gen AIs as collaborators in peer editing activities in a language learning classroom is that they can create an unlimited amount of writing or translation in almost no time. Gen AI-generated texts can also serve as neutral target texts for students to analyze and critique with fewer reservations. This approach fosters a more supportive environment by giving students an alternative source of constructive critique, particularly when peers might feel uncomfortable or unsure about offering corrections. Furthermore, utilizing Gen AIs as both peer translators and reviewers is particularly beneficial in advanced language classes, where students have widely varying levels of language proficiency and educational backgrounds. Gen AIs help level the playing field by accommodating students who need more time to master the material while simultaneously presenting more advanced challenges to those ready for them.

The analysis in section 3 of this paper reveals that the Gen AIs and the three advanced learners of Chinese make similar types of mistakes when their translation outputs are evaluated using the House model. These similarities open up the possibility of using Gen AIs as peers in collaborative learning within advanced Chinese language classes. Since Gen AIs also make these types of mistakes, their output is good material for "mistake-correction" and peer critique activities. It is important to note that Gen AIs do not necessarily produce the same number of mismatches or identical mismatches in the same sentences as human learners. This distinction is central to their pedagogical application as peer learning tools. In practice, as a scaffolding step, the instructor introduces the House model using visual illustrations, such as the flow table shown in Figure 1, and provides guided examples of translation evaluations that illustrate each mismatch type. Students then practice these evaluations using instructor-created materials designed for each mismatch type. Once students become familiar with the House model and can reliably identify mismatch types, they produce their own translations of a given text before generating translations using various Gen AI models. Finally, they compare their translations with those generated by the Gen AIs, critically analyzing differences in nuance, tone, and cultural context based on the criteria in the House model.

For example, one sentence in our Source Text 1 reads: "The tuna belly cake was perfectly prepared with smoky and sweet flavors balanced well by pickled red onions, hot peppers, cucumber, and sprigs of arugula." Students often struggled to choose appropriate expressions for phrases like "perfectly prepared" and "balanced well" and face difficulties breaking down the complex English sentence into coherent Chinese expressions. After students completed their own translations, we showed the Gen AI produced the translation to the students. One student translated it as " 金枪鱼肚饼以烟熏味和甜味完美烹饪，搭配着腌红洋葱，辣椒，黄瓜以及少许芝麻菜 ." ChatGPT-3.5 generated the translation as " 金枪鱼腹蛋糕做得完美，烟熏和甜味恰到好处，腌制的红洋葱、辣椒、黄瓜和芝麻菜的配合相得益彰 ." The comparison highlighted key learning points: the student realized that using " 以……烹饪 " was unnecessary to reflect "prepared…by," and that the long sentence could be broken down into shorter sentences with distinct subjects like " 金枪鱼腹蛋糕 " and " 烟熏和甜味 ." Additionally, expressions

like " 恰到好处 " and " 相得益彰 " were more idiomatic and accurate, matching the Field and Tenor of the original text. Through these comparisons, students identified areas where human input is essential, such as maintaining cultural relevance or refining tone to match the Field, Tenor, and Mode. After the comparison, the researchers further guided the student to critically evaluate the Gen AI's word choice, questioning whether " 蛋糕 " (cake, in Chinese it mainly refers to baked dessert) was a good match for the Field or if " 烟熏和甜味 " (smoky and sweet flavors) was a suitable expression in terms of Mode.

The most valuable gains from the House model in designing and executing collaborative activities that involve Gen AIs is that learners are able to give prompts that will produce more pedagogically meaningful output from the Gen AIs. In peer-critiquing activities, instead of asking generic and general questions such as "how is my usage of words" or "is my grammar correct", students can give prompts informed by the functional mismatch types from House model to the Gen AIs, which evaluates their output. Some Examples of such prompts are given below:

- Do you think the use of passive voice in my Chinese translation fits the attitude of the narrator in the English text? (Syntax-Tenor match)
- Do the word choices in my sentence match the social status of the author, who is an educator? (Lexical-Tenor match)
- The English source text uses second person pronoun "you" to talk to the reader directly, should I use 你 in my Chinese text if I want to achieve the same complex mode of in my translation, or should I omit it? (Lexical-Mode match)

Each functional match or mismatch type corresponds to such prompts depending on the source texts. The responses from the Gen AIs after receiving these more specific prompts contain more useful information for learners. We can adapt such prompts to be used not only in translation activities but in peer-critiquing writing tasks. Students can state their intended functions, instead of the functions of the source texts, and then ask Gen AIs for suggestions regarding their choices of linguistic materials.

4.3 Empowering teachers as facilitators

In language learning, the teacher's role as a facilitator is central to fostering learner autonomy, collaboration, and active learning. According to Brown (2007), facilitative teaching encourages a learner-centered approach, where the teacher provides support and resources while allowing students to take control of their learning processes. Harmer (2007) also notes that a facilitator's role is to provide scaffolding that encourages learners to take risks and experiment with language in a supportive setting. Richards and Rodgers (2014) further highlight that the facilitator approach allows teachers to step back from traditional roles as knowledge transmitters and instead guide learners in exploring language through problem-solving, collaboration, and contextualized use. This shift in roles helps create a learning environment where students feel more responsible for their progress, leading to increased motivation and autonomy.

The discovery from this research empowers the teacher's role as a facilitator House model supplies teachers with questions and prompts that guide students to explore the nuances of the Chinese translation output themselves. For example, when a student translates the sentence "Vianne: Nonsense! Life is too short to deny oneself the pleasure of chocolate. And besides, a little indulgence now and then is good for the soul" as " 维安妮：胡说！人生苦短，及时行乐，而吃巧克力就是最好的乐趣。再说了，偶尔的放纵自己，对身体也有益 ," the student mismatched the medium, using written language instead of spoken language. The teacher can ask guiding questions such as: "What type of text is this? How would you describe the relationship between the two speakers? Are they close? Instead of opting for literary expressions, could informal or spoken language better capture the colloquial and affectionate tone?"

5 Limitations and Conclusion

While the House model provided a valuable framework for our comparative analysis of Gen AI and advanced learner translations, several limitations should be acknowledged. Not all mismatches in the translations in this study fall into the categories laid out in the House model, and one particular type reveals a major challenge to both advanced L2 learners and Gen AIs: the lack or misconception of world/cultural knowledge, which leads to misinterpretation of the source text. One example comes from Text 3, a news report that includes very informal spoken language in direct quotes from bakery workers.

Table 9

Translations of the sentence "And when I see, like, they're taking our photos, 'Oh so pretty,' and then when they eat it, 'Oh, this is so good,' it makes me [feel] like yeah, it's okay," she said. "Like for 15 hours working per day — it's gone."

Translator	Translation
ChatGPT-3.5	她说："当我看到他们拍照时，'哦，好漂亮'，然后当他们吃的时候，'哦，这太好吃了'，我就觉得，嗯，一切都还好。""就像每天工作 15 小时一样——一切都过去了。"
ChatGPT-4o	"当我看到他们拍我们的照片时，说'哦，真漂亮'，然后他们吃的时候又说'哦，这真的很好吃'，这让我感觉，对，这就对了，"她说。"就像每天工作 15 小时——一切都消失了。"
Google Gemini	她说："然后，我看到他们拍照，说'好漂亮啊'，然后吃的时候又说'太好吃了'，这让我觉得，嗯，还行吧。 就像每天工作 15 个小时一样——都消失了。"
Baidu Ernie Bot	"当我看到他们拍照时，说'哦，真好看'，然后当他们吃的时候又说'哦，太好吃了'，这让我觉得'嗯，没关系'，"她说。"就像每天工作 15 个小时——它已经过去了。"
Kimi Moonshot	"当我看到他们，比如，他们拍我们的照片，'哦，这么漂亮，'然后当他们吃的时候，'哦，这太好吃了，'这让我 [感觉] 就像，是的，没关系，"她说。"就像每天工作 15 个小时——都值了。
Student KL	她说："然后我看到他们在拍照，他们说'哇真好看'…然后他们吃的时候就说'哇这个很好吃啊'…我就感觉…是啊…没啥大不了，每天工作 15 个小时一眨眼就过了"。
Student LC	她说道："还有，就是，我看到他们在拍我们的蛋糕时候说：'哇，真好看。' 然后他们吃的时候说：'哇，真好吃。' 我就觉得累一点也没什么。一天十五个小时的辛苦不算什么。'"
Student TL	她说："当我看到他们在拍照时说：'哇好好看啊，'然后他们吃的时候说：'哇这也太好吃了吧'，这让我感觉一切都值得。一天工作 15 小时的疲劳全没了。"

The clause "it's gone" is particularly challenging, with an unknown "it" that cannot be directly retrieved from the text. To understand it, readers need to compare the 15 hours of working time to the regular eight-hour workday and the total of 24 hours in a day. Without this world knowledge, readers may not grasp the implication that the cake-making job is tough and toilsome, and therefore, they may not comprehend the unstated emotional state of the speaker, the "it," or the thing that "is gone" in the sentence. Most human and Gen AI translators did not have or did not consider this crucial information background, and their translations do not convey the intended message of the English source text: the hard feelings caused by the very long working hours are gone because the compliments from clients offer emotional relief and reward. World knowledge and cultural understanding will continue to pose challenges to language learners and, potentially, to Gen AIs, as it takes many years of lived experience and immersion in both cultures to form a comprehensive and interconnected knowledge base.

The House model's benefits for language teaching are most relevant to advanced-level classes. The model focuses on functional matches and mismatches, leaving out more basic forms of errors, such as denotational errors or target language encoding mistakes, which are more common among lower-level language learners. However, we argue that the functional mismatches highlighted by the House model and the general idea of learning a second language as acquiring its linguistic, social, and cultural functions should not be reserved only for advanced learners. It is possible to use the model even with beginning-level learners, especially after they master basic linguistic issues.

The crucial factor in incorporating the House model and Gen AIs into L2 language teaching is the teacher, whose role shifts from traditional lecturing and drilling to negotiating and mediating. Familiarization with the model and differentiation in teaching practice are essential. To address the new challenges posed by Gen AIs and the trend of reconnecting translation with language learning and teaching, teachers need the knowledge and skills to effectively apply the House model.

In conclusion, this study compared the performance of Gen AIs and advanced Chinese learners in translation tasks using the House model of translation quality assessment, revealing that both Gen AIs and advanced learners experienced similar struggles with nuanced cultural and contextual subtleties. Our findings suggest the benefits and feasibility of a collaborative teaching/learning approach, where Gen AIs and advanced learners form peer editing groups for translation and writing revision tasks, assisted by human facilitators. This iterative process of comparison, giving and receiving feedback, and formulating and interpreting prompts fosters independent learning and problem-solving. This collaborative approach not only improves translation skills but also prepares learners to handle the complexities of real-world language use. The House model provides theoretical and practical guidance in setting objectives, selecting teaching materials, designing activities, preparing activities, defining the teacher's role, and conducting formative and summative assessments.

In the future, we will continue testing the linguistic abilities of Gen AI models as they will likely continue to evolve. One project is to compare the revisions of advanced Chinese learners and Gen AIs with extensive prompts drawn from the House model. Another future project is to evaluate the efficacy of this proposed human-AI peer-editing pedagogy, as well as investigate students' subjective responses to their AI partners.

Notes

1. The selected texts include popular novels written within the past 25 years by recognized authors and news articles published on reputable websites within the last 24 months.
2. The three Chinese learners took an HSK mock written test, and all scored 75% or above in the reading section. Their performance on the writing section met the HSK grading criteria for "high-level scores," which require that "the content aligns with the provided materials, the structure is logical, the expression is coherent, and there are no grammatical errors or misspelled characters." Grammar errors in their writing are uncommon, ranging from two to five errors in the 400-character essay. The high proficiency level of the learners in our study is a result of their extensive experience living in a Chinese-speaking environment. The two heritage speakers grew up in large Chinese communities in metropolitan areas where Chinese is used in daily life and business. The non-heritage learner spent extensive time in China.
3. Prompt that ChatGPT-4o received was "could you analyze the genre, tenor, field, and communicative mode of the writing task."

Appendix

Original English Texts and Prompts

English source text	Instructors' prompts
That first bite was divine as were all the others to follow until the plate was clean. The tuna belly cake was perfectly prepared with smoky and sweet flavors balanced well by pickled red onions, hot peppers, cucumber and sprigs of arugula. The herbed yogurt gave a bright, creamy, and cooling effect that rounded out the dish.	This is part of a news article, could you make sure the translation is appealing to the audience and has no authoritative tones.
If you think back to those truffle french fries you were desperate to taste, the first bite may have been amazing, the second bite was probably just as good, and perhaps even the fifth or sixth. But eventually, that feeling of enjoyment begins to diminish.	This is part of a news article with scientific content from a scientist. Could you please make sure the translation delivers the tone?
Dr Rolls reckons that it's an evolutionary tactic that human's developed to keep them healthy. A healthy diet needs to be varied, therefore, our brains have evolved to reward a varied diet: to tell us that changes in what we consume are good.	This is part of a news article with scientific content from a scientist. Could you please make sure the translation delivers the tone?
"And when I see, like, they're taking our photos, 'Oh so pretty,' and then when they eat it, 'Oh, this is so good,' it makes me [feel] like yeah, it's okay," she said. "Like for 15 hours working per day — it's gone."	This is part of a news article, where a cook was explaining her experience. Could you adjust your translation to reflect the tone?
"Instagram can really make people feel like they're not doing enough when they are doing enough," she said. "But when you see something, you know, a little bit messy or a little bit imperfect, it's like, 'Okay, well, I'm a little bit imperfect too, and that looks great. So maybe I'm doing great.'"	This is part of a news article, where a cook was explaining her experience. Could you adjust your translation to reflect the tone?
Jebb told The Times that advertising of junk food was "undermining people's free will" and insisted restrictions were "not about the nanny state". She said: "Advertising means that the businesses with the most money have the biggest influence on people's behaviour. That's not fair.	This is a part of a news article. The speaker is a scientist. Could you please adjust your translation to reflect the tone?
The problem, she said, was to change a "culture in which people, health professionals, doctors, and particularly the sort of powerbrokers in the system, are pretty reluctant to go there". She added: "We can change that. We've changed it with smoking. It took a very long time."	This is a part of a news article. The speaker is a scientist. Could you please adjust your translation to reflect the tone?

I drink a second mouthful, in which I find nothing more than in the first, a third, which gives me rather less than the second. It is time to stop; the potion is losing its magic. It is plain that the object of my quest, the truth, lies not in the cup but in myself. The tea has called up in me, but does not itself understand, and can only repeat indefinitely with a gradual loss of strength, the same testimony; which I, too, cannot interpret, though I hope at least to be able to call upon the tea for it again and to find it there presently, intact and at my disposal, for my final enlightenment.

This is a paragraph from a novel about food and memory. Could you please adjust your translation to reflect the characteristics of this text?

You can let it go in the privacy of your office, you can weep in the walk-in, but at the bench, you must pick up your knife and finish boning out those chickens.

It's possible that working that brunch egg shift at thirty-nine weeks pregnant is badass. And also possible that biting the bullet and scheduling your own labor is badass. Keeping your shit together in front of your crew, no matter what, is badass. Maybe even driving out to IKEA to pick up thirty white china platters and get back by dinner service the day before you are going to give birth is badass.

This is a paragraph from a popular novel featuring the story of a chef starting her own restaurant in NYC. Could you please adjust your translation to reflect the characteristics of the text?

Armande: It's just that... well, I'm trying to watch my waistline. At my age, one must be careful not to indulge too much.

Vianne: Nonsense! Life is too short to deny oneself the pleasure of chocolate. And besides, a little indulgence now and then is good for the soul.

Armande: You always know just what to say, Vianne. Alright, I'll have a small portion of your mousse, but only if you promise to join me.

Vianne: Of course, Armande. I wouldn't dream of enjoying dessert without you.

Armande: Thank you, my dear. Now, tell me, what's your secret to making the perfect chocolate mousse?

This is a conversation in a novel between sisters. Could you please adjust your tone.

Vianne: I'm glad you approve, Armande. There's nothing quite like the pleasure of indulging in a delicious dessert, especially when shared with good company.

Armande: Indeed. And speaking of good company, I'm so grateful to have you here, Vianne. You've brought a bit of sweetness into my life that I didn't know was missing.

This is a conversation in a novel between sisters. Could you please adjust your tone.

"IN MY OPINION, a cake is a lot like a marriage," Lillian began, as she brought eggs, milk, and butter from the refrigerator and put them on the counter. "Admittedly, I don't have a lot of experience," she remarked, holding up her ringless left hand with a wry expression on her face, "but I've often thought that it would be a great idea for couples to make their own wedding cakes, as part of the preparation for their life together. Maybe not so many couples would end up getting married"—Lillian smiled—"but I think those that did might approach it a bit differently."

This is a conversation in a popular novel between a teacher and a student in a culinary school. Could you please adjust your translation accordingly?

References

Aghai, B. (2024). ChatGPT vs. Google Translate: Comparative analysis of translation quality. *Translation Studies Journal*. https://journal.translationstudies.ir/ts/article/view/1156

Baidoo-Anu, M., & Ansah, I. (2023). *Education in the era of generative Artificial Intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning*. http://dx.doi.org/10.2139/ssrn.4337484

Bauermeister, E. (2009). *The school of essential ingredients*. Penguin Books.

Bender, E. M., & Lascarides, A. (2019). *Linguistic fundamentals for natural language processing II: 100 essentials from semantics and pragmatics*. Springer.

Bittner, E. (2020). *Evaluating the evaluator: A novel perspective on translation quality assessment*. Routledge Advances in Translation and Interpreting Studies.

Bosker, B. (2024, March 11). Generative AI and the future of translation: Can technology replace human translators? *The Atlantic*. https://www.theatlantic.com/technology/archive/2024/03/generative-ai-translation-education/677883/

Brown, H. D. (2007). *Teaching by principles: An interactive approach to language pedagogy* (3rd ed.). Pearson Education.

Castilho, S., Mallon, C., Meister, R., & Yue, S. (2023). Do online machine translation systems care for context? What about a GPT model? *Proceedings of the 24th Annual Conference of the European Association for Machine Translation,* 393–417. https://aclanthology.org/2023.eamt-1.39/

Chao, Y. R. (1968). *A grammar of spoken Chinese*. University of California Press.

Cruse, A. (2011). *Meaning in language: An introduction to semantics and pragmatics*. Oxford University Press.

Da Fonte, M. (2023, November 10). The best thing I ate: smoked tuna belly cake from Fyr and Salt. *Northforker*. https://northforker.com/2023/11/the-best-thing-i-ate-smoked-tuna-belly-cake-from-fyr-and-salt/

Dörnyei, Z., & Scott, M. L. (1997). Communication strategies in a second language: Definitions and taxonomies. *Language learning, 47*(1), 173-210.

Drugan, J. (2013). *Quality in professional translation: Assessment and improvement*. Bloomsbury.

Fox, R., Corretjer, O., & Webb, K. (2019). Benefits of foreign language learning and bilingualism: An analysis of published empirical research 2012–2019. *Foreign Language Annals, 52*(4), 699-726.

Gao, Anna (高安娜). (2025). ChatGPT辅助汉语二语写作反馈研究 [A study on ChatGPT-assisted written corrective feedback in teaching Chinese as a second language]. 国际汉语教学学报 *[International Journal of Chinese Language Teaching], 6*(2), 30-43. https://doi.org/10.46451/ijclt.20250203

Goddard, A. (1998). *The language of advertising: Written texts*. Routledge.

Hamilton, G. (2011). *Blood, bones & butter: The inadvertent education of a reluctant chef*. Random House.

Harmer, J. (2007). *The practice of English language teaching* (4th ed.). Pearson Longman.

Harris, J. (1999). *Chocolat*. Doubleday.

He, H., Amini, M., Ganapathy, M., & Che Lah, S. (2024). The application of Juliane House's translation quality assessment model in the English translation of Yu Hua's *Yanre de xiatian. Cogent Arts & Humanities*. https://doi.org/10.1080/23311983.2024.2386781

Hendy, A., Abdelrehim, M., Sharaf, A., Raunak, V., Gabr, M., Matsushita, H., Kim, Y. J., Afify, M., & Awadalla, H. H. (2023). How good are GPT models at machine Translation? A comprehensive evaluation. *arXiv preprint*. https://arxiv.org/abs/2302.09210

Henshaw, F. (2023). Catch my errors if you can: The relative value of peer editing in the language classroom. *NECTFL Review, 90*, 31–41. https://eric.ed.gov/?id=EJ1382637

House, J. (1997). *Translation quality assessment: A model revisited.* Gunter Narr.

House, J. (2014). *Translation quality assessment: Past and present.* Routledge.

Hu, G., & Lam, S. T. E. (2010). Issues of cultural appropriateness and pedagogical efficacy: Exploring peer review in a second language writing class. *Instructional Science, 38*, 371-394. https://doi.org/10.1007/s11251-008-9086-1

Hu, W. (2018). Revisiting translation quality assurance: A comparative analysis of evaluation principles between student translators and the professional trans-editor. *World Journal of Education, 8*(6), 176-186. https://doi.org/10.5430/wje.v8n6p176

Jiang, Z., Xu, Z., Pan, Z., He, J., & Xie, K. (2023). Exploring the role of artificial intelligence in facilitating assessment of writing performance in second language learning. *Languages, 8*(4), 247. https://doi.org/10.3390/languages8040247

Jiang, Z., & Zhang, Z. (2024). Convergences and divergences between automatic assessment and human evaluation: Insights from comparing ChatGPT-generated translation and neural machine translation. *arXiv preprint.* https://doi.org/10.48550/arXiv.2401.05176

Jiao, W., Wang, W., Huang, J. T., Wang, X., Shi, S., & Tu, Z. (2023). Is ChatGPT a good translator? Yes with GPT-4 as the engine. *arXiv preprint.* https://doi.org/10.48550/arXiv.2301.08745

Johnson, A. (2023, August 17). AI could make it less necessary to learn foreign languages. *The Economist.* https://www.economist.com/culture/2023/08/17/ai-could-make-it-less-necessary-to-learn-foreign-languages

Kasper, G., & Kellerman, E. (1997). *Communication strategies: Psycholinguistic and sociolinguistic perspectives.* Routledge.

Kim, J., Yu, S., Detrick, R., & Li, N. (2024). Exploring students' perspectives on generative AI-assisted academic writing. *Education and Information Technologies.* https://doi.org/10.1007/s10639-024-12878-7

Koh, J. (2023, June 20). In New York, a 'cake renaissance' has customers lining up for slices. *The Washington Post.* https://www.washingtonpost.com/food/2023/06/20/new-york-cake-renaissance/

Kohnke, L., Reddy, M., & McCormick, M. (2023). Exploring generative artificial intelligence preparedness among university language instructors: A case study. *Computers and Education: Artificial Intelligence, 5,* 100156. https://doi.org/10.1016/j.caeai.2023.100156

Kroeger, P. (2019). *Analyzing meaning: An introduction to semantics and pragmatics.* Freie Universität Berlin.

Li, J., Ren, X., Jiang, X., & Chen, C. H. (2023). Exploring the use of ChatGPT in Chinese language classrooms. *International Journal of Chinese Language Teaching, 4*(3), 36-55. https://doi.org/10.46451/ijclt.20230303

Li, Q., Zhang, J., & Cai, W. (2024). Utilizing ChatGPT to Implement Differentiated Instruction. *International Journal of Chinese Language Teaching, 5*(1), 74-89. https://doi.org/10.46451/ijclt.20240106

Meng, Kai (孟凯). (2024). ChatGPT 在国际中文词汇教学中的表现与应用策略 [The Performance and Application Strategies of ChatGPT in CSL Vocabulary Teaching]. 国际汉语教学学报 *[International Journal of Chinese Language Teaching], 5*(1), 3-14. https://doi.org/10.46451/ijclt.20240102

Mennella, E., & Quadros-Mennella, D. (2024). Student use, performance, and perceptions of ChatGPT on college writing assignments. *Journal of University Teaching and Learning Practice, 21*(1). https://doi.org/10.53761/pgwk1a93

Min, H. T. (2006). The effects of trained peer review on EFL students' revision types and writing quality. *Journal of Second Language Writing, 15*(2), 118-141. https://doi.org/10.1016/j.jslw.2006.01.003

Nagi, A., Alzain, M., & Naji, T. E. (2023). Informed prompts and improving ChatGPT English to Arabic translation. *Al-Andalus Journal for Humanities & Social Sciences, 11*(98), 211-228.

Pérez-Núñez, A. (2023). Exploring the potential of generative AI (ChatGPT) for foreign language instruction: Applications and challenges. *Hispania, 106*(3), 355-362. https://doi.org/10.1353/hpn.2023.a906568

Proust, M. (2014). *Swann's way* (S. Lee, Ed.; C. Moncrieff Trans.). W.W. Norton & Company.

Pu, Xi (浦曦). (2023). 译者行为批评视角下的ChatGPT翻译质量与其他译本对比研究——以《三体》和《二十大报告》为例 [Comparative Study of Translation Quality between ChatGPT and Other Translations from the Perspective of Translator Behavior Criticism – A Case Study of The Three-Body Problem and Reports of the 20th National Congress]. 现代语言学 [Modern Linguistics], *11*(12): 6107-6113. https://www.hanspub.org/journal/paperinformation?paperid=77670

Richards, J. C., & Rodgers, T. S. (2014). *Approaches and methods in language teaching* (3rd ed.). Cambridge University Press.

Rollinson, P. (2005). Using peer feedback in the ESL writing class. *ELT Journal, 59*(1), 23-30. https://doi.org/10.1093/elt/cci003

Simonsen, A., & Béldi, B. (2023). Using generative AI tools and LARA to create multimodal language learning resources for L2 Icelandic. In *CALL for all languages - EUROCALL 2023 short papers* (pp. 15-18). University of Iceland, Reykjavik. https://doi.org/10.4995/EuroCALL2023.2023.16999

Slade, R. (2023, July 31). Turns out there's a scientific reason that we always have room for dessert. *Delicious*. https://www.delicious.com.au/recipes/florentine-chocolate-tart-coffee-meringue/u46aix0y

Sofyan, R., & Tarigan, B. (2019). Developing a holistic model of translation quality assessment. *Proceedings of the Eleventh Conference on Applied Linguistics (CONAPLIN 2018), 266–271.* Atlantis Press. https://doi.org/10.2991/conaplin-18.2019.267

Song, Fei (宋飞), Tan, Yulu (谭钰璐), & Pu, Yu (浦钰). (2024). 生成式人工智能辅助编写国际中文教育少儿读物的应用与实践——以 ChatGPT与 Midjourney 为例 [Application and Practice of Generative Artificial Intelligence Assisted Writing of Children's Reading Materials in International Chinese Language Education Field]. 国际汉语教学学报 *[International Journal of Chinese Language Teaching], 5*(1), 15-28. https://doi.org/10.46451/ijclt.20240103

Sylvester, R., Smyth, C., & Lay, K. (2023, January 19). Cake in the office should be viewed like passive smoking, obesity expert warns. *The Times*. https://www.thetimes.com/uk/healthcare/article/times-health-commission-office-cake-sugar-passive-smoking-5s3bzb3dn

Tan, S. C., Chen, W., & Chua, B. L. (2023). Leveraging generative artificial intelligence based on large language models for collaborative learning. *Learning: Research and Practice, 9*(2), 125–134. https://doi.org/10.1080/23735082.2023.2258895

Váradi, T. (1980). Strategies of target language learner communication: Message-adjustment. *IRAL: International Review of Applied Linguistics in Language Teaching, 18*(1), 59.

Wang, Li (王力). (1984a). 王力文集第一卷：中国语法理论 *[The complete works of Wang Li, Volume 1, Theory of Chinese grammar]*. 山东教育出版社 [Shandong Education Press].

Wang, Li (王力). (1984a). 王力文集第二卷：中国现代语法 *[The complete works of Wang Li, Volume 2, Modern Chinese grammar]*. 山东教育出版社 [Shandong Education Press].

Williams, M. (2004). *Translation quality assessment: Argumentation-centered approach*. University of Ottawa Press.

Zunaidah, A., Wiharja, C. K., & Febriantono, M. A. (2023). Dependency on AI-based writing tools in English learning: implications for human-computer interaction. *2023 International Conference on Information Management and Technology (ICIMTech)*, Malang, Indonesia, 2023, 1-6. https://doi.org/10.1109/ICIMTech59029.2023.10278054

Dr. Chenqing Song is an Associate Professor of Chinese Studies and Linguistics at Binghamton University, State University of New York, and the coordinator of language programs in the Department of Asian and Asian American Studies. Her main research areas are Chinese phonology, historical linguistics, poetic prosody, and Chinese language teaching. In recent years, she has served as the director of the Translation Research and Instruction doctoral program and has begun exploring the relationship between translation and language teaching. Dr. Song is also a member of the Executive Committee of the International Association of Chinese Linguistics.

Qifei Kao is a Lecturer in Chinese at Binghamton University, State University of New York. Her research interests focus on integrating technology into language classes and applying translation pedagogy to language teaching.

生成式人工智能与高级汉语学习者在翻译任务中的比较研究：基于 House 模型的教学启示

宋晨清
考启菲
纽约州立大学宾汉顿分校，美国

摘要

本研究比较了生成式人工智能（Gen AIs）与高级汉语学习者在翻译任务中的表现差异与共性。研究采用 House（1997, 2014）提出的功能主义翻译质量评估框架，结合话语分析与社会文化视角，从文体类型、语场、语式与语旨四个维度，评估了生成式人工智能与高级学习者所产出的译文的功能对等性。在传统的写作任务中，生成式人工智能凭借庞大的语言数据库，往往能输出语法严谨、结构清晰的句篇；而高级中文学习者则倾向于通过替代表达来弥补语言知识的不足，采用"迂回"或"回避"等交际策略（Váradi，1980；Dörnyei & Scott, 1997；Kasper & Kellerman, 1997）实现表达目的。然而，相较于写作，翻译任务更为严苛，要求译者在不扭曲原意的前提下，精准而得体地传达信息，因此更能考验语言运用的准确性与语境适配能力。本研究中，三位高级汉语学习者和五种生成式人工智能模型分别完成八篇英译中翻译任务（每篇约 400 – 500 英文词）。使用 House 模型进行评估的结果显示，尽管生成式人工智能在语法层面表现出较高的准确性，译文却时常语气生硬、风格失衡或偏离语境；高级汉语学习者的译文虽基本通顺，在语用层面也时有不合语境的表达。鉴于两者错误类型的相似性，本文提出在高级语言教学中引入"协作式修改"教学策略：学习者可在教师提供的话语与功能反馈基础上，对照人工智能生成的译文进行比对与修订，从而深化语言理解与表达能力。本文进一步提出了一套可行性强的教学方案，内容涵盖具体课堂活动与任务设计，旨在为语言教学实践提供新视角与方法参考。

关键词

生成式人工智能，高级汉语学习者，翻译

宋晨清博士现任纽约州立大学宾汉顿分校亚洲研究系终身副教授、语言项目主任。她研究的课题包括汉语音系学、音韵学、历史语言学、诗歌韵律以及对外汉语教学。近年来她担任翻译研究与教学博士项目的主任，把研究方向拓展到翻译与语言教学的关系。宋晨清博士是国际中国语言学会的司库和常务委员会委员。

考启菲现为纽约州立大学宾汉顿分校中文讲师。她的研究兴趣主要为科技与语言教学的融合，以及翻译教学法在语言课堂中的应用。