



## ASSESSMENT OF HUMAN TRANSLATION VS. AI TRANSLATION IN A LITERARY WORK

Doan Thi Thu Phuong\*

*Faculty of English Language and Culture, VNU University of Languages and International Studies,  
No.2 Pham Van Dong, Cau Giay, Hanoi, Vietnam*

Received 13 May 2025

Revised 10 June 2025; Accepted 19 June 2025

**Abstract:** This study investigates the effectiveness of Gemini 2.0 Flash, a large language model developed by Google, in translating humor and stylistic features from David Walliams' *Awful Auntie* into Vietnamese. The AI-generated translation is compared with a human translation by Pham Quoc Hung, using one of Yen Fu's (1854-1921) classical translation principles, stylistic fidelity, as the evaluation framework. Findings indicate that while Gemini 2.0 Flash achieves a degree of literal accuracy, it struggles to convey the original text's humorous tone and stylistic creativity. The AI consistently fails to preserve sound-based rhetorical devices such as assonance and consonance, resulting in a loss of auditory playfulness and comedic effect. It also shows significant limitations in rendering invented lexical items (neologisms), often leaving them untranslated or converting them in rigid, context-insensitive ways. Furthermore, the AI's handling of idiomatic expressions reveals a tendency toward literalism, producing translations that are stylistically unnatural and pragmatically inappropriate. In contrast, the human translator demonstrates flexibility, creativity, and cultural fluency, adapting idiomatic language, wordplay, and stylistic nuance to better suit Vietnamese linguistic and cultural norms. These findings suggest that while AI tools can serve as efficient translation aids, they lack the interpretive depth and creative capacity required for high-quality literary translation, especially in the genre of humorous children's literature. Human translators remain essential in preserving the artistic integrity and expressive impact of such works.

**Keywords:** literary translation, translation assessment, AI translation, human translation, Gemini

---

\* Corresponding author.

Email address: [phuongdoanulis@gmail.com](mailto:phuongdoanulis@gmail.com)

<https://doi.org/10.63023/2525-2445/jfs.ulis.5513>

# ĐÁNH GIÁ BẢN DỊCH CỦA CON NGƯỜI VÀ BẢN DỊCH CỦA AI TRONG MỘT TÁC PHẨM VĂN HỌC

Đoàn Thị Thu Phương

*Khoa Ngôn ngữ và Văn hóa Anh, Trường Đại học Ngoại ngữ, Đại học Quốc gia Hà Nội,  
Số 2 Phạm Văn Đồng, Cầu Giấy, Hà Nội, Việt Nam*

Nhận bài ngày 13 tháng 5 năm 2025

Chỉnh sửa ngày 10 tháng 6 năm 2025; Chấp nhận đăng ngày 19 tháng 6 năm 2025

**Tóm tắt:** Nghiên cứu này xem xét hiệu quả của Gemini 2.0 Flash, một mô hình ngôn ngữ lớn do Google phát triển, trong việc chuyển ngữ các yếu tố ngôn ngữ mang tính hài hước trong tác phẩm *Awful Auntie* của David Walliams sang tiếng Việt. Bản dịch do AI tạo ra được đối chiếu với bản dịch của con người, do Phạm Quốc Hưng thực hiện, dựa trên một trong ba nguyên tắc dịch thuật của Nghiêm Phục (Yen Fu, 1854-1921), cụ thể là độ trung thực về phong cách (stylistic fidelity). Kết quả cho thấy mặc dù Gemini 2.0 Flash có thể tạo ra các bản dịch đúng nghĩa đen ở mức cơ bản, mô hình này gặp nhiều khó khăn trong việc truyền tải giọng văn hài hước và phong cách sáng tạo của nguyên tác. AI thường không duy trì được các yếu tố lấy âm (assonance, consonance), làm mất đi hiệu ứng âm thanh vốn tạo nên tính hài hước trong văn bản gốc. Bên cạnh đó, AI còn gặp hạn chế trong việc xử lý các từ mới do tác giả sáng tạo (neologisms), thường giữ nguyên tiếng Anh hoặc dịch theo cách thiếu linh hoạt. Đáng chú ý, AI cũng không thể truyền tải hiệu quả các thành ngữ (idiomatic expressions), dẫn đến bản dịch nghe cứng nhắc, thiếu tự nhiên hoặc lệch sắc thái. Ngược lại, bản dịch của con người thể hiện sự linh hoạt trong việc truyền tải ý nghĩa theo văn cảnh, sáng tạo trong việc sử dụng từ ngữ phù hợp với văn hóa Việt Nam, và duy trì hiệu quả phong cách hài hước. Nghiên cứu cho thấy rằng mặc dù các công cụ dịch thuật AI có thể hỗ trợ về mặt tốc độ và hiệu suất, chúng hiện vẫn chưa thể thay thế vai trò của người dịch trong việc đảm bảo tính biểu cảm, phong cách và hiệu quả thẩm mỹ của bản dịch văn học, đặc biệt là trong văn học thiếu nhi hài hước.

**Từ khóa:** dịch văn học, đánh giá bản dịch, bản dịch của AI, bản dịch của con người, Gemini

## 1. Introduction

The emergence of artificial intelligence (AI) in translation has sparked considerable debate within the field of translation studies. While AI-powered tools offer substantial advantages in terms of speed, accessibility, and cost-effectiveness, their suitability for translating literary texts, particularly those intended for children, remains a subject of critical scrutiny. Unlike technical or informational content, literary works are inherently rich in metaphor, humor, idiomatic expressions, and culturally embedded references, all of which require more than straightforward linguistic conversion. Literary translation demands a deep engagement with both the source and target cultures, as well as a capacity for creative adaptation that reflects the author's intent and emotional tone.

AI translation systems, despite recent advancements, continue to face several challenges when confronted with the intricacies of literary texts. These include difficulties in interpreting wordplay, puns, allusions, and stylistic devices such as irony or poetic rhythm. Additionally, AI often lacks the contextual awareness necessary to maintain narrative coherence, character voice, or cultural appropriateness, especially when translating for young audiences. Children's literature, in particular, calls for heightened sensitivity to age-appropriate language, humor, cultural references, and educational values. A poorly translated children's book can result not

only in a loss of meaning and entertainment but may also hinder cultural transmission and language acquisition among young readers.

The Vietnamese context further underscores the importance of high-quality literary translation. As English-language children's literature becomes increasingly popular among Vietnamese readers, ensuring the accuracy and cultural resonance of translations is crucial for preserving the educational and moral dimensions of the source text. Translating such literature effectively requires more than literal equivalence; it demands the creative mediation of linguistic play, narrative tone, and socio-cultural context.

This study undertakes a comparative analysis of the translation of David Walliams' *Awful Auntie* and its Vietnamese version, *Bà bác khủng khiếp*, translated by Pham Quoc Hung, to evaluate the effectiveness of AI in literary translation. Specifically, the research examines the capacity of AI - utilizing Gemini Flash 2 as a representative tool - to accurately render linguistic elements while preserving the stylistic and cultural integrity of the source text in the target language. By comparing AI-generated translations with those produced by a professional human translator, this study seeks to assess whether AI can serve as a viable alternative in literary translation or whether it is better suited as a complementary aid in the hands of skilled translators.

## 2. Theoretical Framework and Background

### 2.1. Literary Translation

The concept of literary translation has been defined and interpreted in diverse ways within translation studies, reflecting a range of theoretical orientations and priorities. While scholars agree that literary translation involves more than transferring semantic content from one language to another, they differ in how they frame its purpose, scope, and challenges.

Delabastita (2011) defines literary translation as a practice oriented toward the preservation or recreation of the "aesthetic intentions or effects" of the source text (p. 69). This view emphasizes the translator's responsibility to maintain the stylistic and artistic qualities that characterize literature, including tone, rhythm, wordplay, and narrative voice. From this perspective, literary translation is as much a creative act as a linguistic one, requiring interpretative and expressive skills.

By contrast, Catford (1965), writing from a linguistic perspective, focuses on the structural aspects of translation. He acknowledges that literary translation inevitably involves shifts due to the absence of exact equivalence between source and target languages. For Catford, the challenge lies in managing these shifts without distorting the intended meaning or textual function. While his approach is more technical, it underscores the difficulty of achieving fidelity when dealing with complex literary forms.

Venuti (1995), on the other hand, shifts the discussion from technical and aesthetic concerns to cultural and ideological dimensions. He argues that literary translation is inherently political and shaped by power dynamics between cultures. Venuti introduces the concepts of "domestication" and "foreignization" to describe the translator's strategic choices - whether to make the text familiar to the target audience or to retain its foreign qualities. His approach highlights the translator's visibility and the broader impact of translation on literary reception.

The criteria for a proper literary translation thus vary depending on which perspective is adopted. From Delabastita's standpoint, success lies in recreating the aesthetic experience of the original; for Catford, it involves managing structural differences while preserving meaning;

for Venuti, it entails ethical decision-making about cultural representation. In practice, these criteria intersect and sometimes conflict. A translation that closely mirrors the style and tone of the original may require significant departures from literal meaning, while a linguistically accurate version may lose emotional nuance or cultural depth.

Children's literature adds further complexity to these debates. According to Oittinen (2000), translating for children demands special attention to the target audience's cognitive development, cultural background, and emotional engagement. Wordplay, invented terms, and culturally embedded humor are common in children's books and pose particular challenges for both human and AI translators. Klingberg (1986) argues for "acceptable adaptation" in translating children's literature, stressing that maintaining comprehensibility and relevance is often more important than preserving every original detail.

This study builds upon these contrasting definitions and evaluative criteria by assessing the extent to which artificial intelligence can meet the multidimensional demands of literary translation. Focusing on the English - Vietnamese translation of David Walliams' *Awful Auntie*, the research explores whether AI can reproduce the stylistic, cultural, and audience-sensitive features deemed essential in literary translation theory.

## 2.2. Translation Assessment

Assessing translation quality has long been a focal issue within translation studies, resulting in the development of diverse theoretical frameworks and practical methodologies. Traditionally, early models emphasized linguistic equivalence, focusing primarily on *accuracy* and *fidelity* to the source text (Nida, 1964). These models measured the success of a translation by how closely it replicated the meaning and structure of the original, often prioritizing literalness over cultural or stylistic considerations.

However, such approaches were later critiqued for their limitations in capturing the contextual and cultural dimensions of translation. House (1997), for example, introduced a more nuanced model that incorporated pragmatic and functional parameters, highlighting the necessity of situating translation within its communicative context. This marked a shift toward textual-functionalist paradigms that recognize the translator's role as an intercultural mediator.

A notable development within functionalist theories is the Skopos theory (Vermeer, 1989), which places emphasis on the *purpose* (Skopos) of the translation in the target culture. Here, adequacy is judged by how well the translation meets the needs and expectations of its intended audience, rather than by its closeness to the source text. This represents a fundamental shift from source-oriented to target-oriented evaluation.

In recent years, translation assessment has further expanded to include reader-response and qualitative approaches, acknowledging the inherently subjective nature of translation quality. O'Brien (2011) emphasizes the importance of reader perception, cognitive ergonomics, and usability in evaluating translations, especially in digital and collaborative environments. These developments have broadened the scope of translation assessment beyond linguistic or textual equivalence to include user experience and communicative effectiveness.

Complementing these theoretical models are various practical assessment techniques such as error analysis, comparative analysis, and the use of standardized evaluation metrics. These tools provide systematic methods for identifying issues in translation output and offering constructive feedback to translators, both in professional and academic settings.

While most of these models have emerged from Western academic traditions, this study adopts a historically and culturally distinct approach: the tripartite translation principles of Yen

Fu (1854-1921), a seminal figure in Chinese translation history. In his preface to the Chinese translation of *Evolution and Ethics* by Huxley, Yen Fu outlined three core principles, faithfulness (信, xin), expressiveness (達, da), and elegance (雅, ya), which have profoundly influenced translation theory and practice across East Asia, including in Vietnam.

Yen Fu's model, often referred to as the **Tripartite Translation Paradigm**, emphasizes the following:

**Faithfulness (xin, 信):**

Refers to preserving the essential meaning and intent of the original text. Unlike rigid literalism, Yen Fu advocated conveying the underlying ideas and worldview of the source author.

**Expressiveness (da, 達):**

Prioritizes the intelligibility and clarity of the translated text in the target language. This principle reflects Yen Fu's recognition of the need for the translation to be not only accurate but also accessible to readers from a different cultural and linguistic background.

**Elegance (ya, 雅):**

Pertains to the aesthetic and stylistic refinement of the translated text. For Yen Fu, this often entailed the use of classical Chinese, which signified intellectual sophistication at the time. While more culturally bound, this principle underscores the translator's responsibility to preserve the literary value of the original.

Among these, *ya* (雅) aligns most closely with the modern concept of *stylistic fidelity*, the ability to preserve the tone, register, rhetorical effects, and artistic style of the original text. This includes nuances such as humor, metaphor, irony, or wordplay. In contemporary assessment, stylistic fidelity is crucial for evaluating how well a translation captures the author's creative expression and maintains the intended impact on the reader. Thus, *ya* remains particularly relevant in contexts where the translated text is expected to mirror the literary sophistication or rhetorical flavor of the source material.

Despite being more than a century old, Yen Fu's model continues to offer valuable insights. For this reason, the present study applies his framework, specifically emphasizing stylistic fidelity, to evaluate how well Gemini 2.0 Flash, a machine translation model, renders English texts into Vietnamese compared to human translation. As Fu (1973) noted, "A translation that is faithful but not expressive is no translation at all", reminding us that readability and stylistic elegance are inseparable from the concept of translation quality.

### 2.3. Gemini

Gemini 2.0 Flash represents a significant advancement in Google's suite of large language models (LLMs), building upon the foundation established by its predecessors. Positioned as a "powerful workhorse model" (Google, 2025a), Gemini 2.0 Flash prioritizes efficiency and low latency, making it particularly well-suited for powering interactive and agentic AI experiences.

Key improvements over previous iterations include enhanced performance across various benchmarks, including reasoning, multimodal understanding, mathematical capabilities, and factuality (Google, 2025a). This enhanced performance is coupled with a simplified pricing structure that makes its extended context window, capable of handling up to 1 million tokens, more accessible for developers (Google, 2025b).

The versatility of Gemini 2.0 Flash is evident in its diverse range of applications. Developers are leveraging its capabilities for tasks such as building conversational AI with

natural language understanding, conducting complex data analysis, and creating innovative video editing workflows (Google, 2025b). The model's ability to handle long context windows and provide reliable structured outputs has proven particularly valuable in these applications.

Furthermore, the introduction of Gemini 2.0 Flash-Lite expands the accessibility of this technology. As the most cost-efficient model in the Gemini family, Flash-Lite offers improved performance over earlier versions while providing a more budget-friendly option for projects requiring extensive text output (Google, 2025a).

Overall, Gemini 2.0 Flash represents a significant step forward in the development of LLMs. Its combination of enhanced performance, efficiency, and affordability makes it a compelling tool for developers seeking to create the next generation of AI-powered applications. That is also the reason why the researcher chooses Gemini 2.0 Flash as the representative of AI in comparing AI translation with human translation in this study.

### 3. Methodology

#### 3.1. The book

##### 3.1.1. The Author and the Original English Version *Awful Auntie*

According to <https://www.worldofdavidwalliams.com/about-david/>, David Walliams (real name: David Edward Williams, born August 20, 1971) is a renowned British actor, comedian, and children's book author. Best known for his comedic work on *Little Britain* (2003–2006) and *Come Fly With Me* (2010–2011), Walliams later served as a judge on *Britain's Got Talent* from 2012 to 2022. Since his literary debut in 2008 with HarperCollins, he has authored over 30 titles and sold more than 37 million copies globally. His works have been translated into 53 languages, and his storytelling style has frequently been compared to that of Roald Dahl, blending imaginative adventures with dark humor and emotional depth (World of David Walliams, n.d.).

Among his many successful titles, *Awful Auntie*, Walliams' seventh children's novel, published on 25 September 2014, stands out as both a commercial and critical triumph. The story centers on Stella Saxby, the brave heir to Saxby Hall, who must confront her scheming Aunt Alberta and a fearsome owl named Wagner. With the help of Soot, a ghost who later turns out to be her deceased uncle, Stella thwarts Alberta's attempts to steal her inheritance. Combining humor, horror, and heart, the novel is a signature example of Walliams' ability to engage young readers through outrageous characters and surprising twists.

The choice of *Awful Auntie* for this study is grounded in several compelling reasons. Firstly, the novel has achieved notable literary and cultural impact. It sold over 50,000 copies in its first week, remained at the top of UK children's book charts for seven consecutive weeks, and was the best-selling children's book in the UK in 2014, with over 500,000 copies sold that year. By 2019, it had surpassed 1.2 million copies sold in the UK. It was also the most borrowed children's book in Northern Ireland (2014–2015) and won Audiobook of the Year at the 2014 Specsavers National Book Awards. Its popularity extended to live performance, with the Birmingham Stage Company adapting it into a widely acclaimed theatre production that toured nearly 50 venues with over 500 performances (World of David Walliams, n.d.). Secondly, the novel's linguistic creativity, particularly its use of wordplay, idioms, sarcasm, and humorous exaggeration, makes it an ideal subject for translation studies focused on stylistic fidelity. The challenge of rendering these elements into another language offers fertile ground for evaluating the capabilities of both human and machine translation.

### 3.1.2. The Vietnamese Version: *Bà bác khủng khiếp* - Pham Quoc Hung (Translator)

In Vietnam, the book was published in 2022 under the title *Bà bác khủng khiếp*, translated by Pham Quoc Hung and released by Writers' Association Publishing House. It has received a warm reception from young Vietnamese readers, contributing to the growing popularity of translated children's literature in Vietnam. The availability of a professionally translated Vietnamese version makes the book particularly suitable for comparative analysis in this study, as it allows for direct evaluation of translation strategies employed by a human translator versus those generated by the AI model, Gemini 2.0 Flash. Thus, both the literary significance and stylistic complexity of *Awful Auntie* justify its selection as the primary source text in this research.

### 3.2. Objectives

The primary objective of this study is to conduct a comparative analysis between AI-generated translation and human translation in the context of literary texts. Specifically, the research focuses on evaluating the quality of translation through the lens of stylistic fidelity, defined as the translator's ability to preserve the author's tone, humor, and wordplay. This criterion is particularly critical in children's literature, where narrative voice, rhetorical playfulness, and linguistic creativity are integral to the reader's experience. By examining how these stylistic elements are rendered in both AI and human-translated versions of the same text, the study aims to assess the capabilities and limitations of large language models in handling complex literary language.

### 3.3. Research Instruments and Data Analysis

The corpus for analysis consists of selected excerpts from the original English version of *Awful Auntie* by David Walliams. These excerpts were deliberately chosen based on the presence of stylistically complex features, including consonance (repetition of consonant sounds), assonance (repetition of vowel sounds), neologisms (invented or newly coined words), and idiomatic expressions. These features are hallmarks of Walliams' narrative style and are particularly challenging to translate due to their linguistic playfulness and cultural specificity.

Each selected passage was translated into Vietnamese using the Gemini 2.0 Flash model, Google's high-performance large language model known for its multilingual capacity, extended context window, and low-latency output. The AI-generated translations were then systematically compared to the official Vietnamese version translated by Pham Quoc Hung and published by Writers' Association Publishing House in 2022.

A qualitative content analysis was employed to assess how effectively each version retained the stylistic fidelity of the source text. Particular attention was paid to the preservation or distortion of rhetorical effects such as sound repetition, invented lexical forms, idiomatic usage, and humor. The analysis was guided by Yen Fu's translation principle of elegance (*ya*, 雅), which emphasizes aesthetic and stylistic refinement in translation. Through this framework, the study identifies cases where the translator (human or AI) succeeded or failed to maintain the author's creative intent and literary tone.

## 4. Research Findings

### 4.1. AI Translation vs. Human Translation in Assonance and Consonance

When evaluated on stylistic fidelity, Gemini 2.0 Flash struggled with humor. In order

to create humour in the original English book, David Walliams used a number of wordplay techniques. As a reader who has read both the English origin and the Vietnamese translation by Pham Quoc Hung, the researcher is convinced that he successfully translated David Walliams's wordplay into Vietnamese and still maintained the humour effect, while Gemini 2.0 Flash failed in the majority of cases. Regarding assonance and consonance, Pham Quoc Hung /human obviously outperformed Gemini 2.0 Flash/AI, as illustrated in Table 1.

**Table 1***Assonance and Consonance in Awful Auntie*

No	Assonance and Consonance	Human translation	AI translation
1	Gnash-gnosh	Nghiến ngấu	Nhai nhóp nhép
2	Knicker-knocker-glory	<b>Quần chèn gói</b>	Cầu kỳ lờ loẹt
3	Boom-shack-a-lack	Bắn bùm bùm	Tiếng nổ lớn
4	aunty-waunty	bà bang bác	cô di đáng yêu
5	crashy-washy	bôm bóp bôm bóp	Tiếng đổ vỡ
6	my little niecy-wiecy	cháu bé bông bông bong	cháu gái bé bông
7	illy-willy	ồm yếu yếu yếu	Ngốc nghếch
8	deedsy-weedsies	mấy thứ giấy tờ giấy teo	Việc làm nhỏ nhặt
9	deathy-weathy	ngỏm củ tỏi tỏi toi	Chết chóc đáng sợ
10	Stella-wella-woo-wah	Stella - là - là - la	<b>Stella-wella-woo-wah</b>
11	toilety-woilety (p.127)	vệ sinh vệ seo	nhà vệ sinh
12	poo-poo-poodle-pops	đi bùm bùm bùm	<b>poo-poo-poodle-pops</b>
13	wee-willy-wee-wees	đi xì xì xì	<b>wee-willy-wee-wees</b>
14	owly-wowly	con cú cù cu	<b>owly-wowly</b>
15	thinky-winky	suy nghĩ suy ngheo	suy nghĩ lung tung
16	La-de-da and la-de-day	<b>‘Chanh xả’ và thanh lịch duyên ‘doáng’ quá</b>	Kiểu cách
17	Loopy-loo-loo	Điên rồ điên đại	Lảm nhảm
18	dah-de-dah-de-dum	a bê xê đê	<b>dah-de-dah-de-dum</b>
19	la-de-dah-de-dah	a bê xê đê	<b>la-de-dah-de-dah</b>
20	Itsy-bitsy	xíu xìu xìu	Nhỏ xíu



21	Mincers-mince pies-eyes	Lắt nhắt - vằn với ‘mắt’	Máy xay - bánh bằm - nhìn chăm chăm
22	Hardy-wardy	Khó khăn khăn	Mạnh mẽ
23	Endy-wendy	Rốt cuộc rốt keo	Kết thúc
24	Easy-peasy-poo	Dễ dẽ dề dề	Dễ như ăn kẹo
25	Penny-wenny	Cái bút cái biết	Đồng xu nhỏ
26	Bottom-wottomy	Chỗ dưới chỗ dủng	Mông
27	Girly-wirly	Ngoan ngoãn ngoàn ngoan	Con gái
28	Handy-wandy-pandy	Tay tẩy tẩy tay	<b>Handy-wandy-pandy</b>
29	Especially-wecially	Đặc biệt biên biệt	Đặc biệt
30	Pushy-wushy	Đưa đẩy đẩy đưa	Xô đẩy
31	Pleasy-weasy	Năn nỉ nì nì	Làm ơn

Comparing the assonance, the matching of vowel sounds in language and the consonance, the matching of consonant sounds between the original version and human translation in Table 1, readers can realize that Pham Quoc Hung successfully maintained the majority of assonance and consonance. Human translation only failed to translate assonance and consonance in 2 out of 31 phrases as bold-faced in the column of human translation. Regarding AI translation, Gemini 2.0 Flash could generate the literal meaning of most phrases from English into Vietnamese, but it failed to maintain the assonance and consonance, thus losing the humor effect. In 6 phrases (2, 7, 8, 21, 22, 25), Gemini 2.0 Flash even produced translations that are demonstrably incorrect within the book's narrative, fundamentally altering their intended meaning. For example, in chapter 26 when Aunt Alberta tried to force Stella to sign in the deeds of Saxby Hall, she said “Let me just get you a penny-wenny”; “a penny-wenny” in this context must be a pen (*cái bút cái biết*) as in human translation, and cannot be a coin (*đồng xu*) as in AI translation. AI is also unable to translate several phrases from English into Vietnamese, thus keeping the original English phrases (i.e. *Stella-wella-woo-wah, poo-poo-poodle-pops, wee-willy-wee-wees, owly-wowly, handy-wandy-pandy*). In contrast, human translation is excellent in translating such phrases to create catchy and hilarious wordplay (i.e. *Stella - lả - là - la, đi bum bùm bùm, đi xì xì xì, con cú cù cu, tay tẩy tẩy tay*). In addition, in two phrases 18 and 19, Pham Quoc Hung proved humans’ flexibility and fidelity in translation: “dah-de-dah-de-dum” - a consonance and “la-de-dah-de-dah” - a consonance and assonance in English are both translated into “a bê xê dê” - an assonance in Vietnamese. This phrase is used when a Vietnamese speaker cannot mention all the information about a person, a thing or an event, and the listener can interpret the information in anyway they want. Certainly, AI or Gemini 2.0 Flash is incapable of interpreting this phrase in accordance with Vietnamese culture.

#### 4.2. AI Translation vs. Human Translation in Neologisms

David Walliams invented neologisms - newly coined words or expressions - in his *Awful Auntie*. These unique words and phrases, absent from standard dictionaries and other texts,

posed a significant translation challenge for Gemini 2.0 Flash. This device struggled to translate neologisms in accordance to the contexts they are used in the book. While it could provide basic Vietnamese equivalents, AI failed to replicate the literary style preserved by the human translator, as evidenced in Table 2.

**Table 2**

*David Walliams's Neologisms Translated by Human and AI*

No.	Neologisms	Human translation	AI translation
1	Whipple-scrump	Xoi xềng	<b>Whipple-scrump</b>
2	Winkferno	Xềng lửa	Nháy lửa
3	Knee-thumper	Lên gối	Đập gối
4	Sticky-wink	Xềng keo	Nháy mắt
5	Gigantopot	Chén siêu cao khổng lồ	Chậu khổng lồ
6	Owling, Owlery, Owlcraft, Owlstry, Owlography, Owlosophy	Sành cú, Tổ cú, Nghệ nhân cú, Tuyệt cú, Nhà cú học	Hoạt động xem cú, Nôi nuôi cú, Nghệ thuật cú, Ngành cú học, Nghiên cứu hình ảnh cú, Triết lý về cú
7	owlpert	cú gia	Chuyên gia cú
8	Gonner	Đi tong	Người sắp chết
9	Owlos Hedgehogius	Cú-os Gai-hogius	Cú Nhím
10	Pigowl	Cú lợn	Cú lợn
11	Owlus Smallwingius	Cú-os Cánh-Cụt-ius	Cú Cánh nhỏ
12	Owlus Baldius	Cú-os Trọc-ius	Cú Hói đầu
13	Tripod Owl	Cú Tam Túc	Cú ba chân
14	Voodoodoo Ferns	Dương xỉ tà thuật	Dương xỉ Voodoodoo
15	Cruelberry	Dâu dã man	Quả mọng độc ác
16	Owly?	Cú hơn?	Cú à?
17	Owl-Rack	Bánh răng cú	Giá treo cú
18	Owluettes	Tiểu cú tượng	Cú nhỏ
19	Owleum	Bảo tàng cú	<b>Owleum</b>

Gemini 2.0 Flash could produce the literal or denotative meaning of these neologisms in the majority of phrases; however, its limitations in preserving the original stylistic register render it unsuitable for the translation of children's literature, where reader engagement is

predicated on stylistic fidelity. With two phrases, Gemini 2.0 Flash kept the same English phrases without translation. (i.e. *Whipple-scrump* and *Owleum*), which certainly impede readers' understanding of the translations. In contrast, Pham Quoc Hung even created his own Vietnamese neologisms, the words and phrases which have never been found in any Vietnamese dictionaries (i.e. *Xoi xèng, Xèng keo, Xèng lửa, Cú Tam Túc, Dương xỉ tà thuật, Dâu dã man, Tiểu cú tượng*). The translator is also very creative when he combined English and Vietnamese into the translation of several types of owls (i.e. *Cú-os Gai-hogius, Cú-os Cánh-Cụt-ius, Cú-os Trọc-ius*) in order to maintain the consonance and assonance in the source text.

#### 4.3. AI Translation vs. Human Translation in Idiomatic Expressions

**Table 3**

*Idiomatic Expressions Translated by Human and AI*

No.	Idiomatic expressions	Human translation	AI translation
1.	Her mouth was as dry as a desert.	Miệng cô bé khô như ngói.	Miệng cô ấy khô như sa mạc.
2	Her body felt as heavy as stone.	Cơ thể nặng như chì.	Cơ thể cô ấy nặng như đá.
3.	Lord Saxby managed to keep the bank manager in London at bay	Lãnh chúa Saxby cố gắng xoay sở để <b>lánh mặt</b> vị giám đốc nhà băng giận dữ ở Luân Đôn	Lãnh chúa Saxby đã tìm cách không để người quản lý ngân hàng ở Luân Đôn làm phiền.
4.	Europe was thrust into war.	Chiến tranh nổ ra ở châu Âu.	Châu Âu bị đẩy vào chiến tranh.
5.	Her cooking was atrocious; it became the stuff of legend.	Khả năng bếp núc của mẹ í ẹ, đến mức trở thành <b>giai thoại</b> .	Món ăn cô ấy nấu dở tệ; nó đã trở thành một giai thoại.
<b>6.</b>	<b>Stella wept a river of tears</b>	<b>Stella khóc nức nở.</b>	<b>Stella khóc như mưa.</b>
7.	The girl searched her mind.	Cô bé lục tìm trong ký ức.	Cô gái cố gắng lục lại trong tâm trí mình.
<b>8.</b>	<b>The bird gives me the willies.</b>	<b>Con chim đó khiến tớ phát khiếp đấy.</b>	<b>Con chim làm tôi sợ.</b>
9.	He always kept a stiff upper lip.	Ông luôn cứng cỏi.	Anh ấy luôn giữ thái độ bình tĩnh.
10.	It normally works a treat.	Nó thường có tác dụng.	Nó thường hiệu quả một cách đáng kinh ngạc.
11.	Her mind was racing, and her words are tumbling out.	Đầu óc cô quay cuồng và từ ngữ tuôn ra lộn xộn.	Đầu óc cô ấy quay cuồng, và lời nói cứ tuôn ra không ngừng..
12.	Stella was going to pieces	Stella thấy hoảng loạn.	Stella suy sụp hoàn toàn.

13.	There would be hell to pay.	Đúng là thảm họa	Sẽ có rắc rối lớn.
14.	<b>This whole investigation is rapidly descending into farce.</b>	<b>Toàn bộ cuộc điều tra này đang nhanh chóng trở thành trò hề.</b>	<b>Toàn bộ cuộc điều tra này đang nhanh chóng biến thành trò hề.</b>
15	There's no wriggling out of it.	Đừng có chối nữa.	Không thể chối bỏ được đâu.
16	It was strictly out of bounds to Stella...	Nơi Stella bị cấm bén bằng đến.	Điều đó hoàn toàn bị cấm đối với Stella.
17	<b>Stella had turned the tables on the detective.</b>	<b>Stella đã lật ngược tình thế.</b>	Stella đã lật ngược tình thế đối với thám tử.
18	The detective was sweating like mad	Viên thanh tra vã mồ hôi đầm đìa.	Viên thám tử đổ mồ hôi như điên.
19	She was now on the warpath	Bác ta đang nổi cơn tam bành	Cô ấy giờ đang nổi cơn thịnh nộ.
20	You are completely crackers	Cháu điên thật rồi	Bạn hoàn toàn điên rồ.
21	It was very tough, and not at all tasty, but it was the thought that counted.	Nó rất dai và không ngon một tẹo nào, nhưng của ít lòng nhiều.	Nó rất dai và chẳng ngon chút nào, nhưng quan trọng là tấm lòng.

The data in Table 3 demonstrates striking differences between human and AI-generated translations (Gemini Flash 2) in handling idiomatic and stylistically rich language. These differences are particularly evident when examined through the lens of *stylistic fidelity* - the translator's ability to preserve the tone, voice, rhetorical devices, and expressive nuances of the source text.

In the first two examples, “Her mouth was as dry as a desert” and “Her body felt as heavy as stone,” the human translations - “khô như sỏi” and “nặng như chì” - make use of culturally equivalent idioms that are familiar and vivid to Vietnamese readers. These expressions not only convey the intended meaning but also reflect the original tone and figurative quality. The AI versions - “khô như sa mạc” and “nặng như đá” - are more literal and structurally closer to the source, but they lack the stylistic nuance and naturalness expected in Vietnamese, sounding mechanical and foreign.

A key weakness in AI translations is its difficulty in capturing context-dependent idiomatic meaning. In example 3, the phrase “keep the bank manager at bay” is interpreted by the AI as “không để làm phiền”, which is a direct and surface-level rendering. The human version, “lánh mặt”, is more idiomatic and appropriate in context, conveying the same implication of avoidance with a tone that feels authentic and stylistically aligned with the narrative.

Syntactic restructuring for fluency is another area where the human translators showed a clear advantage. In sentence 4, “Europe was thrust into war,” the human version “Chiến tranh nổ ra ở châu Âu” reads fluently and naturally in Vietnamese, whereas the AI’s “Châu Âu bị đẩy vào chiến tranh” follows the English syntax too closely, resulting in a stilted expression that disrupts the stylistic flow.

Example 5, “Her cooking was atrocious; it became the stuff of legend”, showcases how humor and exaggeration are preserved in the human translation - “Khả năng bếp núc của mẹ í ẹ, đến mức trở thành giai thoại”. This version maintains both the humorous tone and informal register. The AI version, “Món ăn cô ấy nấu dở tệ; nó đã trở thành một giai thoại”, while semantically accurate, is stylistically flat and too formal, missing the tone of the original. Examples 6 to 8 show both versions attempting to use idiomatic equivalents. However, human translations better preserve emotional nuance and colloquial tone. In sentence 8, “The bird gives me the willies”, the human translation “phát khiếp” carries a vivid, emotionally charged reaction, while the AI’s “làm tôi sợ” is neutral and lacks impact.

Cultural awareness also plays a significant role in maintaining stylistic fidelity. In sentence 9, “He always kept a stiff upper lip,” the human version “luôn cứng cỏi” effectively conveys the culturally specific British stoicism. The AI’s “giữ thái độ bình tĩnh” is more generic and lacks the cultural subtext. Similarly, in sentence 10, “works a treat” is modestly rendered as “có tác dụng” in the human translation, capturing the understated tone of the original. The AI’s “hiệu quả một cách đáng kinh ngạc” overstates the effect and disrupts the subtlety of the expression.

In emotionally intense contexts (sentences 11 to 13), the human translations capture psychological nuance more successfully. Phrases like “Her mind was racing” and “going to pieces” are translated as “đầu óc cô quay cuồng” and “hoảng loạn”, respectively, expressions that are concise and emotionally resonant. AI’s versions, while accurate, tend to be longer, overly literal, and less natural in register.

Idioms with strong cultural metaphors such as “out of bounds”, “turned the tables”, or “on the warpath”, are where the human translator excelled. The expressions “bị cấm bển bắng”, “lật ngược tình thế”, and “nổi cơn tam bành” are creative adaptations that preserve both meaning and style. AI translations, though semantically acceptable, fail to carry the same cultural weight or expressive power.

In the final examples (sentences 20 and 21), the gap in stylistic sensitivity becomes even clearer. “You are completely crackers” is translated by the human as “Cháu điên thật rồi”, preserving the informal and humorous tone. AI’s version, “Bạn hoàn toàn điên rồ”, feels stiff and unnatural. The idiom “It’s the thought that counts” is adeptly rendered by the human as “của ít lòng nhiều”, a culturally equivalent proverb that reflects the source’s intent and tone. In contrast, the AI’s “quan trọng là tấm lòng” is correct but sounds like a literal interpretation rather than an idiomatic expression.

Overall, while Gemini Flash 2 can produce semantically accurate translations, it fails to preserve the stylistic dimensions of the source text. Literalism, lack of cultural awareness, and rigid syntactic transfer result in translations that are technically correct but stylistically off-tone. The human translator, by contrast, consistently demonstrates the ability to adapt not just meaning but mood, rhythm, and cultural resonance, ensuring that the target text reads fluently and retains the expressive impact of the original. This analysis reaffirmed the continuing superiority of human translation in preserving stylistic fidelity, especially in literary and idiom-rich contexts.

## 5. Discussions

The findings of this study highlighted a marked discrepancy in translation performance between human translators and artificial intelligence (AI) systems, particularly in terms of *stylistic fidelity*. Through the comparative analysis of selected linguistic features in *Awful*

*Auntie* - including assonance, consonance, neologisms, and idiomatic expressions - it is evident that human translation demonstrated significantly greater sensitivity to tone, creativity, and target-language norms.

The analysis of assonance and consonance revealed the limitations of AI in reproducing stylistic sound patterns that contribute to the playful rhythm and musicality characteristic of David Walliams's narrative style. While the AI model, Gemini 2.0 Flash, was able to preserve basic denotative meaning, it consistently failed to replicate the intended rhetorical effects created by sound devices. The absence of attention to alliteration, internal rhyme, and phonological cohesion in the AI-generated output resulted in translations that lacked the auditory appeal necessary for engaging young readers. In contrast, the human translator effectively restructured the source text in Vietnamese to maintain both the stylistic liveliness and phonetic parallelism of the original.

With regard to neologisms, the findings further underscored the creative limitations of AI. Translating invented words requires more than lexical substitution; it demands an intuitive grasp of tone, genre, and cultural expectations. The AI model tended to either transliterate unfamiliar terms or retain English neologisms without adapting them to Vietnamese morphological or phonetic norms, thereby disrupting the reader's immersion and coherence of the target text. By contrast, the human translator displayed considerable linguistic dexterity, coining inventive equivalents that preserved the playful spirit and imaginative tone of the original while aligning naturally with Vietnamese linguistic structures.

Idiomatic expressions, as another key stylistic element, posed a consistent challenge for AI-generated translations. While the AI model was generally competent in recognizing the literal meaning of idioms, it often failed to convey their pragmatic function or emotional tone. As a result, many translations appeared literal, formal, or culturally incongruent, diminishing the expressiveness of the text. In contrast, the human translator demonstrated a strong command of idiomatic Vietnamese, skillfully adapting figurative language in a way that resonated with the target audience and preserved the author's intended humor, irony, or emphasis. This capacity to balance semantic content with stylistic function is essential in literary translation and remains beyond the reach of current AI capabilities.

These findings have several implications for both machine translation development and translator training. From a technological standpoint, the inability of AI systems to reproduce stylistic features such as sound play, lexical creativity, and idiomatic nuance indicates that human oversight remains indispensable, particularly in genres like children's literature where tone and engagement are paramount. Current AI tools may offer useful first drafts or assist with literal content rendering, but they lack the cultural and stylistic intuition required for fully polished literary translation.

From a pedagogical perspective, the study underscores the need to prioritize stylistic fidelity as a core competency in translator education. Future translators must be equipped not only with bilingual proficiency but also with the ability to analyze tone, identify register-specific idioms, and creatively adapt expressive features across languages. Training activities that focus on re-creating neologisms, preserving poetic rhythm, and interpreting idioms in context can help students develop the interpretive and creative skills necessary for high-quality literary translation. Moreover, exposure to AI-generated output in classroom settings can offer students valuable opportunities to critically assess stylistic shortcomings and refine translations accordingly.

In conclusion, while advances in neural machine translation continue to enhance lexical

and syntactic accuracy, this study reaffirmed that stylistic fidelity, particularly in idiomatic, figurative, and creative language, remains a uniquely human strength. The human translator's role as a cultural mediator and stylistic interpreter is not only relevant but essential in preserving the voice, humor, and artistic integrity of the original literary work.

## 6. Conclusion and Recommendations

This study has offered a critical comparison between human and AI-generated translations in the context of literary texts, with a particular focus on *stylistic fidelity*, a core evaluative criterion derived from Yen Fu's Tripartite Translation Paradigm. By examining the Vietnamese translation of David Walliams' *Awful Auntie*, the research reveals that while the AI model Gemini 2.0 Flash performs adequately in rendering literal meaning, it consistently fails to preserve stylistic nuances that are central to the effectiveness and emotional resonance of children's literature.

Across the three examined stylistic features, assonance and consonance, neologisms, and idiomatic expressions, the AI model exhibited notable limitations. It struggled to replicate the phonological creativity and rhythmic charm of Walliams' prose, failed to invent or adapt culturally coherent neologisms, and often produced idiomatic translations that lacked pragmatic relevance or tonal appropriateness. In contrast, the human translator, Pham Quoc Hung, displayed a high degree of linguistic inventiveness and cultural fluency, successfully adapting stylistic elements in a way that preserved the author's voice and humor for Vietnamese readers.

These findings reaffirm the limitations of current AI systems in translating literary texts, especially in genres such as children's literature where tone, rhythm, and creativity are essential to reader engagement. While AI tools like Gemini Flash 2 offer speed, consistency, and semantic accuracy, they remain constrained by their lack of cultural awareness, contextual reasoning, and stylistic sensitivity. Consequently, their role should be considered complementary, useful for generating initial drafts or handling surface-level content, but not yet capable of replacing human translators in contexts requiring expressive depth and aesthetic nuance.

The study makes both theoretical and pedagogical contributions. Theoretically, it affirms the ongoing relevance of stylistic fidelity, aligned with Yen Fu's principle of *elegance* (雅), as a robust evaluative criterion in literary translation. Pedagogically, it emphasizes the importance of training translators not only in bilingual competence but also in literary adaptation, creative language use, and the interpretation of tone, humor, and figurative language. Translator education programs should incorporate comparative translation exercises, post-editing of AI-generated texts, and stylistic rewriting tasks to help students identify and correct deficiencies in machine translation.

However, this study is not without limitations. Firstly, its scope was limited to one literary work *Awful Auntie* and therefore might not be fully generalizable to other genres or authors. The findings, while indicative, were context-specific and centered on a single children's text rich in humor and wordplay. Secondly, only one AI system (Gemini 2.0 Flash) and one human translation were compared; future studies could incorporate a broader range of AI models and human translators to allow for more representative comparisons. Thirdly, the analysis focused exclusively on stylistic fidelity, excluding the two other principles in Yen Fu's tripartite model, *faithfulness* and *expressiveness*. As such, the assessment remained partial and would benefit from a more holistic evaluation framework in future research. Finally, this study did not empirically measure reader response or reception, which could offer additional insights into how AI and human translations were perceived in practice.

These limitations point to several directions for future inquiry. Subsequent research could expand the comparative framework by including multiple literary texts from different genres, apply all three of Yen Fu's criteria, and incorporate empirical data from target readers to assess reception and comprehension. Additionally, as AI continues to evolve, future studies should monitor improvements in machine learning models trained on stylistically complex corpora, and explore hybrid translation workflows that combine AI efficiency with human stylistic intervention.

In conclusion, while AI-powered tools such as Gemini Flash 2 represent significant technological progress, their current limitations prevent them from achieving stylistic equivalence in literary translation. The nuances of humor, idiomatic richness, and rhetorical flair remain beyond the grasp of even the most advanced models. As such, human translators continue to play an irreplaceable role - not merely as linguistic conveyors but as creative cultural mediators - in preserving the integrity, beauty, and resonance of literary texts across languages and cultures.

### References

- Catford, J. C. (1965). *A linguistic theory of translation*. Oxford University Press.
- Delabastita, D. (2011). Literary translation. In Y. Gambier & L. van Doorslaer (Eds.), *Handbook of Spanish translation studies* (pp. 69–78). John Benjamins. <https://doi.org/10.1075/9789027203328>
- Fu, Y. (1973). General remarks on translation (C. Y. Hsu, Trans.). *Renditions*, 1, 4–6.
- Google. (2025a, February 5). *Gemini 2.0: Flash, Flash-Lite and Pro*. Google Developers Blog. <https://developers.googleblog.com/en/gemini-2-family-expands/>
- Google. (2025b, February 25). *Start building with Gemini 2.0 Flash and Flash-Lite*. Google Developers Blog. <https://developers.googleblog.com/en/start-building-with-the-gemini-2-0-flash-family/>
- House, J. (1997). *Translation quality assessment: A model revisited*. Gunter Narr Verlag.
- Klingberg, G. (1986). *Children's fiction in the hands of the translators*. Almqvist & Wiksell International.
- Nida, E. A. (1964). *Towards a science of translating: With special reference to principles and procedures involved in Bible translating*. Brill.
- O'Brien, S. (2011). Collaborative translation. In Y. Gambier & L. van Doorslaer (Eds.), *Handbook of translation studies: Volume 2* (pp. 17–20). John Benjamins. <https://doi.org/10.1075/hts.2.col1>
- Oittinen, R. (2000). *Translating for children*. Garland Publishing.
- Venuti, L. (1995). *The translator's invisibility: A history of translation*. Routledge.
- Vermeer, H. J. (1989). *Skopos and commission in translational action*. Oy Finn Lectura Ab.