



ARTICLE

<https://doi.org/10.1057/s41599-025-05379-6>

OPEN

 Check for updates

# Normalization or creation? A corpus-based study of normalization in the Chinese translation of English children's literature

Yang Han<sup>1</sup>✉

Based on the English-Chinese Parallel Corpus of Children's Literature and the Corpus of Chinese Children's Literature, this study investigates the feature of normalization in the Chinese translation of English children's literature. Normalization refers to the adaptation of foreign features in the source text to comply with the cultural and linguistic norms of the target culture. The study analyzes both macro and micro levels of language features in translated children's literature, comparing them with original Chinese and English texts. The findings reveal a clear trend towards normalization, evidenced by shorter sentences, increased repetition of high-frequency words, a lower frequency of hapax legomena, and a higher textual readability in translated Chinese versions. Furthermore, linguistic structures such as reduplication, modal particles, “把” (BA), and “得” (DE) constructions are found to occur at rates comparable to or significantly higher than those in the original Chinese corpus. This paper argues that normalization is a creative outcome, molded by translators aligning with reader expectations, conscientiously considering the psychological characteristics of child readers, and adapting to social, cultural, and market influences. The study contributes to understanding linguistic features of translated children's literature, sheds light on translation universals, and underscores the dynamic interplay between normalization and translator creativity.

<sup>1</sup>University of Shanghai for Science and Technology, Shanghai, China. ✉email: [hanyang@usst.edu.cn](mailto:hanyang@usst.edu.cn)

## Introduction

The normalization of translated language, or standardization, refers to “the tendency of translators to adapt foreign features of the source text to the cultural and linguistic inventory in the target culture” (Tully, 2014, p.295). Rather than simply aligning the translated text with the linguistic conventions of the target language, normalization is a multifaceted process that reflects the translator’s strategic decisions, the genre of the text, and the intended communicative purpose. This process encompasses a wide range of linguistic elements, including punctuation, vocabulary selection, sentence structure, style, and discourse structure (Vanderauwera, 1985, p. 93). It can extend to the exaggerated use of target language norms (Baker, 1996, p. 183). Examples of such standardization include incorporating incomplete sentences, creative vocabulary, collocations, or punctuation that conforms to the standardized target language (May, 1997; Kenny, 2001; Du et al. 2021; Santos, 2024). Additionally, it involves adopting typical expressions and usages of the target language (Ke, 2003; Hu, 2006; Pan, 2014; Wu, 2019; Wu and Li, 2021), as well as integrating cultural normalization features (Hu et al. 2021). In some cases, translators may amplify certain target language features to achieve specific effects, such as emphasizing readability and fluency.

While related concepts such as simplification and explication are often discussed in translation studies, these terms refer to distinct processes. Simplification, for example, involves the reduction of complexity in the source text to make the target text more accessible, often by simplifying grammatical structures, vocabulary, or the overall message. Unlike normalization, simplification does not inherently focus on aligning with or exaggerating the conventions of the target language, but rather seeks to make the content easier to understand without adhering to specific linguistic norms (Baker, 1996). Explication, on the other hand, involves making implicit information in the source text more explicit in the target text, such as adding clarifications, connectives, or explanatory notes (Blum-Kulka, 1986). Explication is a communicative strategy aimed at ensuring clarity and unambiguity, distinct from normalization, which focuses more on aligning with target language norms and conventions.

The concept of normalization has often been viewed as a translation universal, implying that it is an inherent feature of the translation process (Baker, 1993, p. 246) and is not influenced by specific languages (Laviosa, 1998, p. 288). However, recent debates have questioned the universality of this concept. Scholars such as De Sutter et al. (2012) and Jia et al. (2022) argue that the notion of translation universals may be more hypothetical than universal. Chesterman (2010) suggests a more nuanced approach to understanding translation universals, proposing that the term “universal” be used in a “weaker” sense to refer to statistical patterns rather than absolute traits of translations. Furthermore, Chesterman (2014) advocates for a more context-dependent analysis, taking into account factors such as language pair, translator profiles, working conditions, and the specific genre of the text.

Given these debates surrounding normalization as a potential translation universal, it becomes evident that normalization is not merely a matter of conforming to target language norms. Rather, it is a complex process shaped by the interaction of various factors, including translation strategies, cultural considerations, and the translator’s goals—factors that align with Chesterman’s (2014) call for a context-dependent analysis. This suggests that translation tendencies should be examined within specific linguistic, cultural, and situational frameworks.

Currently, scholarly inquiries into the normalization features of literary works predominantly center on adult literature, leaving a noticeable gap in exploring children’s literature. Empirical studies

on the language of translated children’s literature have primarily focused on simplification and Europeanization in comparison to original Chinese children’s literature (Puurtinen, 2003; Yu, 2014; Zang, 2010). Other studies have delved into the language of translations of children’s literature from a reader-centric perspective. For instance, Wang (2016), based on an empirical investigation of Chinese child readers, demonstrated that the identifiability of language in translated Chinese children’s literature is relatively low, and the disparity from the language of original Chinese children’s literature is not substantial. This study provides a general overview of the tendency for translated language to resemble that of the original works. However, to achieve a more thorough understanding of the specific application of typical language structures in translated Chinese children’s literature, it is imperative to conduct empirical investigations using extensive corpora.

This study investigates normalization features and influencing factors in the English-to-Chinese translation of children’s literature at both macro and micro levels, by comparing translated texts with original English texts and contemporary Chinese children’s literature. It aims to test the validity of the normalization hypothesis in this genre while also delving into the characteristics and factors that impact the translation of children’s literature into Chinese. Through this analysis, the research contributes to a deeper understanding of how language, ideology, and literary conventions shape translation practices in cross-cultural contexts.

## Materials and Methods

This article is based on a self-compiled corpus of contemporary Chinese translations of English children’s literature, specifically works translated or published since 2001. This time span is selected based on relevant studies indicating a significant rise in the translation and publication of foreign children’s books in China. According to research by Wan (2019), China’s children’s literature publishing entered a period of diversified development starting in 2001. Notably, the number of translated children’s books increased significantly after this period, reflecting broader access to international works. This turning point coincided with China’s accession to the World Trade Organization, which facilitated the expansion of the publishing industry and intensified cultural exchange. Given these developments, the timeframe provides an appropriate foundation for analyzing normalization features and the underlying factors shaping translation practices.

The Corpus consists of two sub-corpora: the English-Chinese Parallel Corpus of Children’s Literature (Corpus A) and the Corpus of Chinese Original Children’s Literature (Corpus B) (hereinafter referred to as the “Translation Corpus” and the “Chinese Original Corpus,” respectively).

Corpus A includes translations by distinguished translators, selected based on their professional credentials, awards, and recognition within the translation community, and published by reputable publishing houses specializing in high-quality children’s literature. These works have been chosen as required reading or classic reading lists in China’s new curriculum standards issued by the Ministry of Education of the People’s Republic of China<sup>1</sup>. The selection of these works also considers their high sales and positive reviews on e-commerce platforms. For instance, data from Dangdang’s official bestseller rankings<sup>2</sup> show that several of these translated works consistently rank among the top-selling children’s books, reflecting their popularity and acceptance among readers, which is particularly relevant for understanding market-driven translation practices. On Dangdang.com, these works have collectively received well over 2,000,000 reviews, with

**Table 1 Composition of the Corpus.**

Sub-Corpus	Title	Author/Translator	Publisher	Year
the English-Chinese Parallel Corpus of Children's Literature (Corpus A)	<i>The Lion, the Witch and the Wardrobe</i>	Trans. Chen Liangting, Liu Wenlan	Yilin Press	2022
	<i>Peter Pan</i>	Trans. Yang Jingyuan	Central Compilation & Translation Press	2020
	<i>Charlotte's Web</i>	Trans. Ren Rongrong	Shanghai Translation Publishing House	2014
	<i>The Witches</i>	Trans. Ren Rongrong	Tomorrow Publishing House	2009
	<i>Harry Potter and the Philosopher's Stone</i>	Trans. Cao Suling, Ma Ainong	People's Literature Publishing House	2019
	<i>The Wind in the Willows</i>	Trans. Ren Rongrong	Shanghai Translation Publishing House	2012
the Corpus of Chinese Original Children's Literature (Corpus B)	<i>Jia Li, the Boy &amp; Jia Mei, the Girl</i>	Qin Wenjun	Zhejiang Juvenile & Children's Publishing House	2018
	<i>Bronze and Sunflower</i>	Cao Wenxuan	Jiangsu Juvenile & Children's Publishing House	2005
	<i>Laughing Cat Diary: The Secret Paradise of Children &amp; Laughing Cat Diary: The Blue Rabbit's Ear Grass</i>	Yang Hongying	Tomorrow Publishing House	2018
	<i>The Little Green Man</i>	Ban Ma	21st Century Publishing House	2011
	<i>Three Magic Dragons</i>	Peng Yi	JieLi Publishing House	2012
	<i>My Mom is an Elf</i>	Chen Danyan	Juvenile & Children's Publishing House	2006

**Table 2 Size of the Corpus.**

Sub-corpus	English	Chinese
A	306,956	461,954
B		432,178
Total	306,956	894,132

an average satisfaction rate exceeding 99%, according to publicly available sales data and user ratings on the platform<sup>3</sup>.

Corpus B comprises works by well-known domestic children's literature authors, selected based on their significant contributions to the field, recognition by literary awards, and influence on young readers in China. The themes of the works are also consistent with Corpus A. The selection of these authors and their works ensures that the corpus represents the broader landscape of Chinese children's literature and allows for a meaningful comparison with the translated works in Corpus A.

The detailed list of works included in both sub-corpora is presented in Table 1 below.

The detailed size of each corpus is presented in Table 2.

According to Table 2, the size of the two sub-corpora is very similar, with 461,954 characters in sub-corpus A and 432,178 characters in sub-corpus B. The time span for both sub-corpora, encompassing works since 2001, allows for a focused analysis of the recent trends and developments in the translation of children's literature into Chinese. Moreover, All the selected works in both corpora belong to the genre of novels and children's literature. And works in Corpus A such as "*The Lion, the Witch and the Wardrobe*" and "*Harry Potter and the Philosopher's Stone*", as well as those in Corpus B like "*The Little Green Man*" "*Three Magic Dragons*" and "*My Mom is an Elf*" share the characteristics of fantasy literature. The themes of the two corpora are also consistent, covering growth and self-discovery (such as "*Charlotte's Web*", "*Peter Pan*" from Corpus A, and "*Jia Li, the Boy*", "*Jia Mei, the Girl*" from Corpus B), courage and adventure (such as "*The Lion, the Witch and the Wardrobe*", "*Harry Potter and the Philosopher's Stone*" from Corpus A, and "*Three Magic Dragons*" from Corpus B), the struggle between good and evil (such as "*The*

*Witches*" from Corpus A and "*My Mom is an Elf*" from Corpus B), family and friendship (such as "*The Wind in the Willows*" from Corpus A and "*Bronze and Sunflower*" from Corpus B). Additionally, all the works share a recommended age range for readers, specifically targeting the 7–12 years old. This similarity in terms of corpus size, temporal scope, genre, theme, and recommended age group establishes a notably favorable basis for comparability between the two sub-corpora. Given the slight difference in corpus size, all frequency counts of linguistic features are presented with both raw numbers and normalized frequencies (per 10,000 characters) to ensure rigorous and meaningful cross-corpus comparison.

At the overall level, this study investigates several key variables, including average sentence length, average clause length, high-frequency words, hapax legomena, and text readability. Average sentence length denotes the average number of characters in complete sentences, while average clause length refers to the average number of characters in sentence segments. Given the developmental characteristics of children's age and language proficiency, children's literature typically features "short sentences" (Jiang, 2013, p. 17). If the average sentence length and average clause length of the translation corpus are equal to or smaller than those of the Chinese Original Corpus, it can be considered indicative of the normalization features inherent in the translated works. To determine the average sentence length and average clause length, periods, exclamation marks, and question marks were identified as sentence markers, while commas, semicolons, colons, along with periods, exclamation marks, and question marks, were designated as clause markers. The software AntConc 3.5.9 (Anthony, 2020) was employed to compute the relevant metrics, which were subsequently compared across different corpora. High-frequency words are those that appear frequently within the corpus, whereas hapax legomena are words that occur only once. The utilization of high-frequency words and the presence of hapax legomena serve as indicators, as observed in studies by Wang and Hu (2008), Hu (2010), and others, to assess the normalization of texts. In general, a higher frequency of high-frequency words and a lower occurrence of hapax legomena in a text suggest reduced creativity and heightened normalization in the language. The last variable at the

overall level is text readability, which refers to the ease with which a reader can understand a written text. It is influenced by factors such as vocabulary diversity, sentence complexity, and clarity of meaning. In this study, readability is closely related to normalization as it reflects the adjustments made by translators to align translated texts with the linguistic and cognitive expectations of the target audience. These adjustments often involve simplifying vocabulary, refining sentence structures, and ensuring semantic clarity, all of which contribute to the normalization process. To assess text readability, this study employed AlphaReadabilityChinese (Lei et al. 2024), a computational tool developed for evaluating the readability of Chinese texts across three linguistic dimensions—lexical, syntactic, and semantic—based on nine quantitative indicators. Unlike earlier tools that focused primarily on surface-level features such as word and sentence length, AlphaReadabilityChinese incorporates more sophisticated and robust algorithms that enable a nuanced analysis of textual complexity. While the tool was initially validated using literary texts by Jin Yong and Gu Long, its applicability has been recognized across multiple disciplines, including digital humanities, and international Chinese language education. Although the tool has not been specifically tailored to translated children's literature, its design enables it to capture general trends in language use—such as lexical variation, syntactic diversity, and semantic clarity—that are relevant to the investigation of normalization in translation. To mitigate limitations inherent in automated analysis, this study supplements the readability metrics with micro-level manual analyses of distinctive grammatical features (e.g., reduplication, modal particles, “把” and “得” constructions), thereby ensuring a more comprehensive evaluation of stylistic tendencies in translated versus original texts.

At the specific word and sentence level, the variables examined include reduplication, modal particles, “把” (BA), and “得” (DE) constructions, which are all typical features of the Chinese language. Reduplication, also known as repetition or doubling of characters, is formed by repeating the same syllable, morpheme, or word, and can be classified into continuous or intermittent reduplication. Modal particles are particles used to express tone or emotion at the end of or pause within a sentence, with commonly used particles including “啊” (a), “呀” (ya), “吗” (ma), and others. This study first examines the application of these two types of words in the Chinese Original Corpus. If the frequency of these structures in the translation corpus is significantly higher than that in the Chinese Original Corpus or there is no significant difference, there is a tendency of normalization of the translated text. “把” (BA) and “得” (DE) constructions are typical sentence structures in Chinese. The “把” (BA) construction introduces the recipient of the verb action before the predicate verb and is mostly used to indicate a disposal meaning. The “得” (DE) construction, on the other hand, introduces the complement after the predicate (verb or adjective) and is used to describe or evaluate the action or quality of the predicate. The usage frequency and sentence length of these two types of constructions were quantitatively retrieved from both corpora. In addition to this quantitative comparison, representative contextual instances were examined in the discussion section to explore how these constructions contribute to text accessibility and emotional expression in children's literature. A tendency toward normalization is suggested if the frequency in translated texts is significantly higher or comparable to that in original texts, and if the sentence structures are similarly concise and aligned with conventional syntactic patterns in Chinese. This combined analysis considers not only structural features but also how such structures enhance readability and support child reader comprehension in specific narrative contexts.

## Results

**Overall Level.** At the overall level, the investigated variables include average sentence length, average clause length, high-frequency words, hapax legomena, and text readability. Periods, exclamation marks, and question marks were used as sentence markers, while commas, semicolons, colons, periods, exclamation marks, and question marks were designated as clause markers. The data were processed using AntConc 3.5.9 software, and the results of the calculations are presented in Table 3.

As depicted in Table 3, the translation corpus exhibits a slightly shorter average sentence length (19.67) in comparison to the Chinese original corpus (21.7). This shorter sentence length in translations reflects a strategic choice by translators to enhance readability and ensure clarity for a younger audience, which is characteristic of children's literature. The average clause length of the two corpora is quite similar, with a minor difference (8.65 vs. 8.18), which suggests that clause structure remains relatively consistent across the original and translated texts.

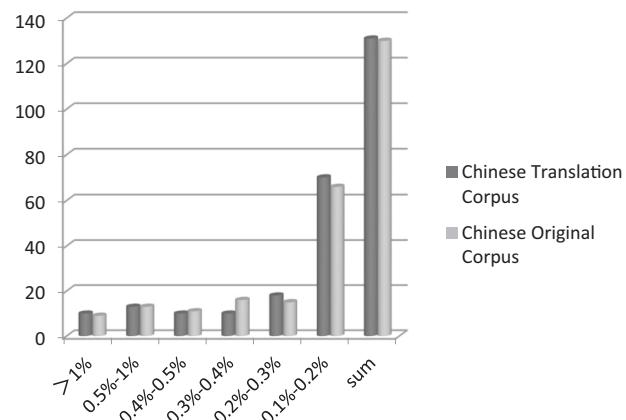
The frequency and proportion of high-frequency words and hapax legomena were analyzed using AntConc 3.5.9 to generate word lists for both corpora. The detailed findings are presented in the following sections.

The initial analysis focused on the distribution of high-frequency word types across various frequency bands (1% or more, 1–0.5%, 0.5–0.4%, 0.4–0.3%, 0.3–0.2%, 0.2–0.1%), as depicted in Fig. 1.

As seen in Fig. 1, the overall number of high-frequency word types with a frequency ratio of 0.1% or higher in the translation corpus is marginally greater than that in the Chinese Original Corpus. Specifically, across each frequency range, excluding the 0.3–0.4% range where the number is lower, the number of high-frequency word types in the other ranges is either comparable to or slightly higher than that in the Chinese Original Corpus. This finding reflects linguistic adaptation strategies intended to align more closely with the target language's stylistic norms for children's literature.

**Table 3 Comparison of the Overall Characteristics of the Chinese Translation Corpus and the Chinese Original Corpus.**

Overall Characteristics	Chinese Translation Corpus	Chinese Original Corpus
Average Sentence Length	19.67	21.70
Average Clause Length	8.65	8.18



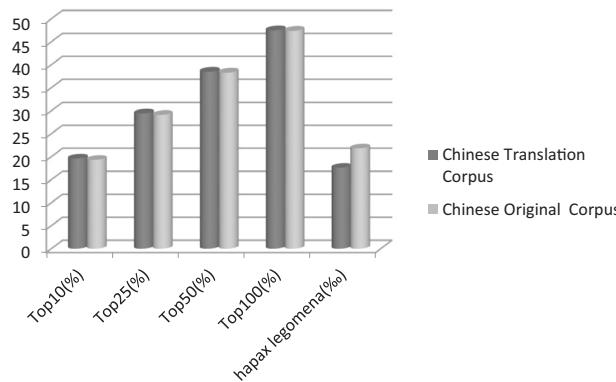
**Fig. 1** Comparison of the Number of High-frequency Words in the Chinese Translation Corpus and the Chinese Original Corpus.

Furthermore, Fig. 2 illustrates the proportions of high-frequency words ranked in the top 10 (e.g., 的, 了, 他, 我, 是, 在, 说), top 25 (e.g., 他们, 她, 地, 就, 上, 这, 它, 把), top 50 (e.g., 什么, 没有, 想, 好, 看, 走), and top 100 (e.g., 吧, 给, 吗, 还, 像, 起来), along with the proportion of hapax legomena (e.g., 哀悼, 哀叫, 哀伤, 哀怨, 挨饿, 挨揍).

According to Fig. 2, the translation corpus has slightly higher proportions of high-frequency words ranked in the top 10, top 25, top 50, and top 100 compared to the Chinese Original Corpus. Conversely, the proportion of hapax legomena is notably lower in the translation corpus than in the Chinese Original Corpus. The statistical analysis reveals that the total number of hapax legomena in the translation corpus is 5528, whereas in the Chinese Original Corpus, it is 6510. The difference in the frequency of hapax legomena in the Chinese Translation Corpus is statistically significant ( $p < 0.01$ ), which points to a tendency in translations to avoid the use of highly unique or rare words. This suggests that translators prioritize certain familiar, common words in their translations to make the text more accessible to younger readers.

The following section presents a comparison of the text readability between the Chinese Translation Corpus and the Chinese Original Corpus, as outlined in Table 4. The table provides two primary metrics for readability: overall readability and readability standard deviation. The overall readability scores offer an indication of the general accessibility of the corpus, while the readability standard deviation reflects the variation in readability across individual texts within each corpus.

The comparison of text readability between the Chinese Translation Corpus and the Chinese Original Corpus, as shown in Table 4, provides insights into the linguistic normalization characteristics of translated texts. The analysis considers various



**Fig. 2** Comparison of the Frequency of High-frequency Words and Hapax Legomena between the Chinese Translation Corpus and the Chinese Original Corpus.

readability metrics, including lexical richness, syntactic richness, semantic accuracy, semantic richness, semantic clarity, and semantic noise. For lexical richness, the Chinese Original Corpus exhibits a higher overall value (6.54) than the Chinese Translation Corpus (6.19). This suggests that the original texts display greater lexical variability, potentially increasing their reading difficulty. In contrast, the lower standard deviation in the Chinese Translation Corpus (0.29 vs. 0.39) indicates a more consistent use of vocabulary, reflecting a higher degree of linguistic normalization across the translated texts. The syntactic richness values are closely aligned between the two corpora (2.23 vs. 2.25), with slightly higher variability observed in the Chinese Original Corpus (0.02 vs. 0.01). This minimal difference suggests that both corpora maintain comparable syntactic complexity, though the translated texts exhibit greater structural uniformity, contributing to enhanced readability. In terms of semantic accuracy, the Chinese Translation Corpus generally demonstrates higher values across multiple dimensions. For instance, the semantic accuracy for nouns (4.62 vs. 4.23), noun-verb combinations (7.82 vs. 7.47), and content words (7.84 vs. 7.59) are all higher in the translation corpus, indicating simpler and more accessible word choices. The reduced standard deviations in the translation corpus further highlight its internal consistency in semantic accuracy, contrasting with the greater variability observed in the original texts. The semantic richness of the Chinese Translation Corpus (0.11) is notably lower than that of the Chinese Original Corpus (0.15), implying less topical diversity in the translated texts. Similarly, the semantic clarity is slightly reduced in the translations (0.03 vs. 0.04), though both values are low, reflecting a focus on concentrated and straightforward thematic expression in both corpora. A significant disparity is observed in semantic noise, with the Chinese Translation Corpus showing a markedly lower value (278.27) compared to the Chinese Original Corpus (699.92). This reduction suggests that the translated texts are less prone to irrelevant or unimportant topics, contributing to improved textual coherence and focus. Moreover, the analysis of standard deviation values across these metrics reveals a pronounced trend of lower variability in the Chinese Translation Corpus. This indicates a higher degree of internal uniformity, suggesting that the language used in the translated texts adheres more closely to linguistic norms.

In summary, children's literature translated into Chinese exhibits a slightly shorter average sentence length than Chinese original children's literature, while the average clause length remains relatively similar. The frequency of high-frequency words is relatively high, whereas the frequency of hapax legomena is significantly lower. The readability statistics reveal a tendency toward greater linguistic consistency and reduced lexical variation in the translated texts. Lower lexical richness and reduced semantic noise suggest that translated children's literature

**Table 4 Comparison of the Text Readability of Chinese Translation Corpus and Chinese Original Corpus.**

Text Readability	Overall Readability		Readability Standard Deviation	
	Corpus	Chinese Translation Corpus	Chinese Original Corpus	Chinese Translation Corpus
lexical_richness	6.19	6.54	0.29	0.39
syntactic_richness	2.23	2.25	0.01	0.02
semantic_accuracy_n	4.62	4.23	0.27	0.59
semantic_accuracy_v	9.53	9.67	0.42	0.48
semantic_accuracy_n_v	7.82	7.47	0.33	0.45
semantic_accuracy_c	7.84	7.59	0.28	0.32
semantic_richness_n	0.11	0.15	0.01	0.02
semantic_clarity_n	0.03	0.04	0.03	0.02
semantic_noise_n	278.27	699.92	115.28	171.10

adheres more closely to standardized linguistic norms, with fewer unexpected or diverse word choices. Additionally, the lower variability across readability metrics indicates a more homogeneous discourse style compared to the original texts. Together, these findings point to a pattern of linguistic normalization in translated children's literature, which is characterized by more uniform vocabulary, greater textual coherence, and a reduced degree of lexical and topical diversity.

**Specific Level.** At the specific word and sentence level, the variables investigated include reduplication, modal particles, “把” (BA) and “得” (DE) constructions. The software AntConc 3.5.9 was employed to retrieve and quantify the usage frequency and sentence length of these variables across both corpora.

**Reduplication.** Reduplication can be retrieved using regular expressions to search for continuous and intermittent reduplicated words in the corpus. The retrieved reduplicated words are then manually reviewed, screened, and categorized. The extracted reduplicated words include AA (e.g., 看看 “look-look”, 慢慢 “slow-slow”, 紧紧 “tight-tight”; extended forms like AAA克克克 “ke-ke-ke”, AAAA 哈哈哈哈哈 “ha-ha-ha-ha”), AABB (e.g., 结结巴巴 “stammer-stammer”, 清清楚楚 “clear-clear”), and ABAB (e.g., 嘎吱嘎吱 “creak-creak”, 咔哒咔哒 “click-clack”) patterns for complete reduplication, as well as ABB (e.g., 冷冰冰 “cold-icy”), ABAC (e.g., 自言自语 “talk to oneself”), and ABCB (e.g., 一动不动 “not move at all”) patterns for partial reduplication. The usage patterns of different types of reduplicated words in the Chinese Original Corpus were analyzed first, as presented in Table 5.

According to Table 5, the high-frequency ( $> 10\%$ ) reduplicated words in the Chinese Original Corpus are AA (including AAA and AAAA), ABB, AABB, and ABAC patterns. A comparative frequency analysis was conducted for these four types of reduplicated words in the translation corpus, with the detailed results shown in Table 6.

Based on the above table, the frequency of the four types of reduplicated words in the translation corpus is relatively similar to that in the Chinese Original Corpus. The results of the log-

likelihood ratio test (sig) indicate that there is no significant difference in the frequency of these four types of reduplicated words between the two corpora ( $p > 0.05$ ). Hence, the prevalence of the four high-frequency reduplicated word types in the translation corpus mirrors that in the Chinese Original Corpus, indicating a normalization trend in classic Chinese translations of children's literature.

**Modal Particles.** Following segmentation and POS tagging of the corpus texts using ICTCLAS (Zhang and Liu, 2002), modal particles in both corpora were identified through manual verification, and their frequencies were calculated for each corpus. To ensure a systematic comparison, the frequencies and statistical significance of all identified modal particles were analyzed across both corpora. The results are shown below.

Based on Table 7, a total of 16 types of modal particles were identified across both corpora. In the Chinese Translation Corpus, modal particles appear 2792 times (60.44 per 10k), significantly more than in the Chinese Original Corpus (42.97 per 10k) ( $p < 0.01$ ). The most frequent modal particles in the Chinese Original Corpus (with occurrences over 100) are “啊” (4.54 per 10k), “吧” (8.86 per 10k), “吗” (10.62 per 10k), “呢” (10.57 per 10k), and “呀” (2.89 per 10k), all of which also occur frequently in the translation corpus. Among these, “啊” “吧” “吗” and “呢” appear significantly more frequently in the translation corpus ( $p < 0.01$ ), with occurrences of 434 (9.39 per 10k), 731 (15.82 per 10k), 712 (15.41 per 10k), and 556 (12.04 per 10k), respectively. Although the frequency of “呀” in the translation corpus (2.4 per 10k) is slightly lower than that in the Chinese Original Corpus (2.89 per 10k), the difference is not statistically significant ( $p > 0.05$ ). As for less frequent particles, most do not show significant differences between the two corpora ( $p > 0.05$ ), with only a few, such as “呵” and “哇”, showing marginally significant differences ( $p < 0.05$ ). However, these particles are rare in children's literature and do not meaningfully affect the overall trend.

Overall, the inclusion and comparison of all identifiable modal particles reveals a statistically significant increase in modal particle usage in the translated texts ( $p < 0.01$ ). This finding supports the tendency toward normalization in children's literature translation.

**“把” (BA) sentences.** The character “把” (BA) was used as a search term to extract sentences containing it from both corpora. After manual proofreading to exclude sentences where “把” was not used as a “BA” structure, the frequency of “把” (BA) sentences was calculated in each corpus. The results are shown in the following table.

According to Table 8, the Chinese Original Corpus contains 942 instances of “把” (BA) sentences, while the translation corpus contains 1723 instances. The results of the log-likelihood ratio test indicate that the frequency of “把” (BA) sentences in the translation corpus is significantly higher than that in the Chinese Original Corpus ( $p < 0.01$ ).

**Table 5 Frequency of Reduplicated Words in the Chinese Original Corpus.**

Reduplication	Frequency	Percentage (%)
AA, AAA, AAAA	2496	56.869
ABB	504	11.483
AABB	458	10.435
ABAB	286	6.516
ABAC	567	12.919
ABCB	78	1.777
Total	4389	100.000

**Table 6 Comparison of the Frequency of Reduplicated Words Between the Chinese Translation Corpus and the Chinese Original Corpus.**

Reduplication	Chinese Original Corpus		Chinese Translation Corpus		Significance of Difference	
	Raw frequency	Normalized frequency (per 10k)	Raw frequency	Normalized frequency (per 10k)	Loglikelihood	Sig.
AA, AAA, AAAA	2496	57.75	2567	55.57	1.88	0.17
ABB	504	11.66	477	10.33	3.63	0.057
AABB	458	10.60	433	9.37	3.36	0.067
ABAC	567	13.12	584	12.64	0.4	0.529

**Table 7 Comparison of the Frequency of Modal Particles Between Chinese Translation Corpus and Chinese Original Corpus.**

Modal Particles	Chinese Original Corpus		Chinese Translation Corpus		Significance of Difference	
	Raw frequency	Normalized frequency (per 10k)	Raw frequency	Normalized frequency (per 10k)	Loglikelihood	Sig.
啊(a)	196	4.54	434	9.39	77.02	0 *** -
吧(ba)	383	8.86	731	15.82	88.6	0 *** -
呗(bai)	0	0.00	2	0.04	\	\ \ -
呵(he)	26	0.60	2	0.04	26.04	0 *** +
啦(la)	94	2.18	130	2.81	3.66	0.056 -
哩(lí)	4	0.09	10	0.22	2.27	0.132 -
喽(lou)	4	0.09	5	0.11	0.05	0.815 -
吗(ma)	459	10.62	712	15.41	39.54	0 *** -
嘛(ma)	44	1.02	47	1.02	0	0.997 +
么(me)	20	0.46	0	0.00	\	\ \ +
哪(na)	23	0.53	33	0.71	1.19	0.275 -
呐(na)	4	0.09	13	0.28	4.44	0.035 * -
呢(ne)	457	10.57	556	12.04	4.22	0.04 * -
哇(wa)	6	0.14	1	0.02	4.3	0.038 * +
噢(o)	8	0.19	4	0.09	1.64	0.2 +
哦(o)	4	0.09	0	0.00	\	\ \ +
呀(ya)	125	2.89	111	2.40	2.03	0.155 +
哟(yo)	0	0.00	1	0.02	\	\ \ -
Total	1,857	42.97	2,792	60.44	132.2	0 *** -

**Table 8 Comparison of the Frequency of “把” (BA) Sentences in the Chinese Translation Corpus and the Chinese Original Corpus.**

“把” (BA) Sentences	Chinese Original Corpus		Chinese Translation Corpus		Significance of Difference	
	Raw frequency	Normalized frequency (per 10k)	Raw frequency	Normalized frequency (per 10k)	Loglikelihood	Sig.
Total	942	21.80	1723	37.30	183.19	0.000

**Table 9 The Length of “把” (BA) sentences in the Chinese Translation and Original Corpora.**

the Length of “把” (BA) Sentences	Chinese Original Corpus		Chinese Translation Corpus		Significance of Difference	
	Frequency	Percentage (%)	Frequency	Percentage (%)	Loglikelihood	Sig.
≥30	0	0.000	4	0.200	\	\ \ -
25-29	4	0.425	4	0.200	0.72	0.397 +
20-24	8	0.849	27	1.600	2.56	0.109 -
15-19	47	4.989	101	5.900	0.85	0.357 -
10-14	255	27.070	376	21.800	6.95	0.008 ** +
5-9	511	54.246	929	53.900	0.01	0.912 +
1-4	117	12.420	282	16.400	6.52	0.011 * -
Total	942	100.000	1723	100.000	0.00	1.000 1 0

The length of the structure was measured following the method outlined by Hu and Zeng (2011), by calculating the number of characters between the character “把” (BA) and the first punctuation mark that follows it in both the translation corpus and the Chinese Original Corpus. The lengths of “把” (BA) sentences in both corpora are presented in Table 9.

According to the above table, the length of “把” (BA) sentences in the Chinese Original Corpus is similar to that in the translation corpus, with the 5-9 range having the largest proportion, followed by the 10-14 range, and the 1-4 range. Judging from the log-likelihood ratio test results of the frequency of “把” (BA) sentences in each range in the two corpora, in 1-4 range, the “把” (BA) sentences in the Chinese Translation Corpus accounted for relatively more than that in the Chinese Original Corpus ( $p < 0.05$ ). In the 10-14 range, the proportion of “把” (BA)

sentences in the Chinese Original Corpus is significantly higher than that in the translation corpus ( $p < 0.01$ ). In the 5-9 range and other length categories, the frequency of “把” (BA) sentences in the two corpora is identical, showing no significant difference ( $p > 0.05$ ). Accordingly, the lengths of “把” (BA) sentences in the two corpora are relatively similar, and they do not show significant differences like the non-children's literary corpus (According to Hu and Zeng (2011), the number of longer “把” (BA) sentences in the translation corpus is higher than that in the Chinese Original Corpus, while the number of “把” (BA) sentences with few words in the translation corpus is lower than that in the Chinese Original Corpus).

From the above analysis, it can be seen that the frequency of typical Chinese “把” (BA) sentences in the Chinese Translation Corpus is significantly higher than that in the Chinese Original

**Table 10 Comparison of the Frequency of “得” (DE) Sentences in the Chinese Translation Corpus and the Chinese Original Corpus.**

“得” (DE) Sentences	Chinese Original Corpus		Chinese Translation Corpus		Significance of Difference	
	Raw frequency	Normalized frequency (per 10k)	Raw frequency	Normalized frequency (per 10k)	Loglikelihood	Sig.
Total	978	22.63	1,056	22.86	0.05	0.82

**Table 11 Length of “得” (DE) Sentences in the Chinese Translation Corpus and the Chinese Original Corpus.**

Length	Chinese Original Corpus		Chinese Translation Corpus		Significance of Difference	
	Frequency	Percentage (%)	Frequency	Percentage (%)	Loglikelihood	Sig.
≥25	1	0.102	0	0.000	\	\
20-24	0	0.000	1	0.095	\	\
15-19	8	0.818	14	1.326	1.23	0.268
10-14	29	2.965	49	4.640	3.77	0.052
5-9	206	21.063	181	17.140	4.10	0.043 *
2-4	659	67.382	722	68.371	0.07	0.787
1	75	7.669	89	8.428	0.36	0.546
Total	978	100.000	1056	100.000	0.00	1.000 1 0

Corpus, and the length of “把” (BA) sentences is similar. It shows that the Chinese translation of children’s literature tends to use the typical Chinese sentence pattern, and the length of “把” (BA) sentences has not been significantly expanded, showing a trend consistent with original Chinese children’s literature, reflecting the normalization of the language of children’s literature translation.

“得” (DE) sentences. The character “得” (DE) was used as a search term to extract sentences containing it from both corpora. Non-“得” (DE) sentences were manually reviewed and excluded, and the frequency of “得” (DE) sentences was then calculated for each corpus. The results are shown below.

Table 10 shows that there are 978 sentences in the Chinese Original Corpus and 1056 sentences in the Chinese Translation Corpus. The results of the log-likelihood ratio test show that there is no significant difference in the frequency of “得” (DE) sentences in the two corpora ( $p > 0.05$ ), indicating that the number of “得” (DE) sentences in the translation corpus is close to that in the Chinese Original Corpus.

The structural length of “得” (DE) sentences was also examined in both corpora. Following the same statistical method applied to “把” (BA) sentences, the number of characters between “得” (DE) and the first punctuation mark was counted. The statistical results of the length of “得” (DE) sentences in the translation and original corpora are shown in Table 11.

Based on Table 11, further analysis of the sentence structure of “得” (DE) sentences in both corpora revealed that the majority of “得” (DE) sentences had a length of 2-4 characters, accounting for nearly 70% in both corpora. The second most common length range was 5-9 characters, accounting for approximately 20% in both corpora, followed by 1-character length sentences, which accounted for approximately 8% in both corpora. Long sentences (over 20 characters) containing “得” (DE) were rare in both corpora, with the longest sentence containing 26 characters in the Chinese Original Corpus and 23 characters in the translation corpus. Overall, the statistical analysis showed that the sentence length distribution of “得” (DE) sentences in both corpora was similar, with short sentences (1-9 characters) being the most common.

In summary, compared to the original Chinese children’s literature corpus, the translated Chinese children’s literature

corpus exhibits a normalization trend, with a slightly shorter average sentence length, a higher frequency of high-frequency words, fewer hapax legomena, slightly improved text readability, a similar form and frequency of reduplicated words, a greater number of commonly used modal particles, a higher quantity of “把” (BA) constructions with similar structural lengths, a comparable number of “得” (DE) constructions, and both “把” (BA) and “得” (DE) sentences being predominantly short in form.

## Discussion

For children’s books, there’s an expectation that authors use language skillfully, reflecting their own culture’s grace while keeping it understandable for children (Cao, 2014, p. 47). Translators of children’s literature share this responsibility, sometimes emphasizing certain unique features. Based on the above findings of the study, it becomes apparent that translated children’s literature showcases normalization features both at the overall level and in specific linguistic aspects.

Regarding the reasons for normalization, some argue that these features are inherent in the translation process itself, considering them universal due to the unavoidable constraints present in translation (Baker, 1993, p. 246). Others propose that normalization is a strategic outcome, driven by the translator’s desire to mitigate risks and to align the translated work with the cultural norms of the target language. Pym (2008) contends that literal translations serve as a safe option, especially when faced with ambiguities in determining the exact meaning of the source text. However, such translations may lead to awkward or unnatural phrasing in the target language. To avoid this, translators often turn to high-frequency forms or other normalized expressions in the target language, ensuring both naturalness and effective communication with a broader readership. This paper argues that the normalization features identified in translated children’s literature reflect the creative decisions made by translators to meet the expectations of young readers.

**Expectations of child readers.** The normalization features observed in translated classics of children’s literature are intricately linked to the expectations of child readers in the target

language. In line with Chesterman's (1997) development of Toury's norm theory, this study draws attention to the role of expectancy norms in shaping translation decisions. Expectancy norms pertain to readers' expectations regarding their language proficiency, esthetic preferences, and cultural conventions of the target language. These norms play a pivotal role in ensuring that translations are accessible and engaging to the intended audience.

In the realm of translating children's literature, the designated audience is explicitly identified as children, typically ranging from ages 7 to 12. Significantly, there exists a notable age and esthetic disparity between translators and readers. To ensure the translation is well-received by the intended young audience, translators commonly employ a child-centric approach. This involves taking into account the language acquisition characteristics and esthetic psychology of child readers.

Firstly, the normalization features observed in translated children's literature are closely linked to children's language acquisition stages, particularly during the primary school years (ages 7–12). According to experimental studies, children process words differently from adults due to their developing cognitive and linguistic skills. For example, word frequency significantly influences children's word-processing efficiency. Similar to adults, children take longer to process low-frequency words compared to high-frequency ones, which can impact their overall reading fluency (Blythe et al., 2009). In addition, word length is another critical factor affecting children's word processing. Studies have shown that children find it easier to process shorter words than longer ones, as shorter words reduce the cognitive load during reading (Joseph et al., 2009; Li, 2021). These support the translators' tendency to favor high-frequency words, simplify sentence structures and prefer shorter ones when adapting complex source texts for child readers. Such adaptations ensure that young readers can follow the narrative more effortlessly, thereby enhancing their reading experience.

Moreover, at this stage, children are undergoing critical development in both mechanical and semantic memory, which involves internalizing language structures and expanding their vocabulary. It is widely recognized that standardization plays a crucial role in this process (Li, 1986). Li (2016, p. 108) emphasizes that the chosen pieces in primary school Chinese textbooks should demonstrate standardized language and elegant expression. Children's literature holds a crucial position as the main reading material during this formative stage. The normalization aspect in these works contributes to fostering children's use of standardized language and is deemed essential for their understanding and appreciation of the literature. The focus on high-frequency words and shorter sentence structures in translated texts can be seen as an effort to accommodate the cognitive and linguistic needs of young readers. The normalization observed at the overall level, including the use of high-frequency words and short, simple sentences, appears to align with pedagogical principles for language acquisition. Additionally, sentences and clauses are intentionally kept short, underscoring the emphasis on brevity and appropriate punctuation. The translation does not exhibit a significant influence from the English source text, avoiding a Euro-centric syntactic feature. For instance, sentence 1 "*Poor Mole found it difficult to get any words out between the upheavals of his chest that followed one upon another so quickly and held back speech and choked it as it came*" is translated as "可怜的鼹鼠的胸口一下一下起伏得太快了, 话刚要出口就被呛下去, 觉得很难说出话来" (*The poor mole's chest rose and fell too quickly, words before spoken were choked back, feeling very difficult to speak a word*). This translation simplifies the original structure by breaking it into three shorter sentences, reflecting a clear intention to make the text more accessible to younger readers. The above sentence can also be translated literally as "可

怜的鼹鼠发现在胸口那一阵阵快速接踵而至的起伏之间很难把任何话说出来, 刚要说出口的话被阻止了"(The poor mole found that between the rapid and successive upheavals in his chest, it is difficult to speak any words, words before spoken were prevented). Both the literal translation and the normalized translation are grammatically correct in Chinese. The literal translation, though closely mirroring the original English structure, still maintains coherence in Chinese syntax, though it may sound slightly more formal or less fluid for young readers. The normalized version, on the other hand, adjusts the sentence structure to align with the common practices of Chinese children's literature writing style, ensuring clarity and readability without distorting the meaning. The use of shorter sentences, natural punctuation, and simpler constructions makes it more suitable for children's reading comprehension. At the specific level, the frequent use of "把" (BA) and "得" (DE) constructions in translations demonstrates a strategic adaptation that aligns with both Chinese syntactic norms and child reader expectations. For example, sentence 2 "*they (the beds) can be hidden behind the screens on first Thursdays*" is translated as "头几个礼拜四, 可以把床藏在屏风后面" ("In the first few Thursdays, can put the beds hidden behind the screens"), converting a passive clause into a direct and vivid dispositional action. This transformation facilitates comprehension and narrative engagement. Similarly, sentence 3 "*He had a strange, but pleasant little face*" is translated as "他那张小脸长得怪怪的, 但很愉快" ("His little face looks a bit strange but is very pleasant"), where the use of "得" not only conveys evaluation but also introduces emotional tone and rhythm, aiding children's affective understanding. These adaptations at the specific level demonstrate a strategic choice in maintaining linguistic features conducive to the target audience's language development and understanding while adhering to Chinese language norms.

Regarding the impact of the translations of the above sentences on young readers, we conducted a controlled reader feedback survey with 42 primary school children (aged 7–12) from three classes in a Shanghai urban school. The participants were selected through stratified random sampling, ensuring representation across different grade levels (Grade 4 – Grade 6). The survey employed a paired-stimulus design where participants were presented with eight pairs of translated sentences (including literal vs. normalized versions) and asked to select the version they found easier to understand, more natural in speech, more vivid, or more in line with their reading expectations (see Appendix for full questionnaire).

The results of the survey reveal a clear preference for the normalized version. A majority of the respondents (more than 86% for sentence 1, 74% for sentence 2, and 60% for sentence 3) indicated that the normalized translation was easier to understand and more aligned with their reading expectations. Chi-square tests indicated that these differences were statistically significant ( $p < 0.05$ ). These findings suggest that simplified sentence structures, natural punctuation, and child-friendly linguistic choices play a crucial role in enhancing young readers' comprehension and engagement with translated texts.

Secondly, child readers have high esthetic expectations for the expressive qualities of language, which differ from those of adult readers. Concrete, tangible elements such as shapes, colors, sounds, and images are more likely to capture their attention (Zhang, 2010, p. 256). Expressive, colloquial, and rhythmic language is more conducive to engaging children, helping them connect emotionally with the text. The incorporation of typical Chinese words and phrases in translated children's literature serves to enhance the expressiveness of the language. Reduplication, for example, involving characters with the same initial consonants, vowels, and tones (sometimes changing), is

recognized for its rhythmic and poetic characteristics in Chinese (Huang, 1993, p. 27). Frequent reduplicated expressions like “呆呆” (dull dull), “抽抽搭搭” (choppy choppy or jerky jerky), “毛茸茸” (furry furry), “嘎吱嘎吱” (creaky creaky), “咚咚咚” (thump thump thump), compared to non-reduplicated forms make the text more colloquial and enhance the sensory expressiveness of sounds and images. Exclamatory particles like “呢” (ne) at the end of sentences convey the speaker's tone and emotion, contributing to the colloquial and interactive characteristics of the translation, and helping distinguish a character's internal thoughts from third-person narration, thereby enriching the reading experience. For example, the particle “呢” appears in the following two sentences:

- (1) 这又有什么关系呢? 现在什么都无所谓了。(What again does it matter **ne**? Now, nothing matters anymore.)  
The use of “呢” here marks the sentence as an inner rhetorical question, reinforcing the character's emotional detachment and sense of hopelessness.
- (2) 这会儿, 那些东西正在阿斯兰身上爬来爬去呢。(At this moment, those things are crawling on and off Aslan **ne**.)

In this sentence, “呢” indicates the continuity and immediacy of the action, echoing the character's internal observation and anxiety.

These instances of “呢” serve to reflect the character's inner thoughts, distinguishing them from the surrounding narrative and vividly portraying the character's emotional state and perspective.

Additionally, “把” (BA) constructions often highlight subjectivity (Hu, 2009, p. 111–115) and focus attention on the outcome, thereby emphasizing emotions and tone. For instance, the sentence “Did you really manage to pour the whole bottle into their soup?” is translated as “你当真把整瓶东西都倒到她们的汤里去了吗?” (“You really put the whole bottle of things all poured into their soup?”). In this case, the translation employs the “把” (BA) construction combined with the intensifier “都” (DOU) to highlight the outcome of the event. This choice emphasizes the speaker's surprised tone, making the language more dynamic and fitting for children, compared to a direct translation like “你真的成功倒了整瓶东西到他们的汤里吗?” (You really succeed in pouring the entire bottle of things into their soup?). Finally, “得” (DE) constructions are often used in literary Chinese to create a colloquial tone. (Zhang, 2000, p. 105–108). For example, the sentence “They are perfectly safe, aren't they? Every one of the little angels sound asleep in bed” is translated as “他们都安全得很,是不是? 三个小天使都在床上睡得正香呢” (“They all safe very much, right? The three little angels all in bed sleep very well.”). Compared to a literal translation like “很安全” (very safe) and “熟睡” (sound asleep), the translation with “得” (DE) better aligns with Chinese colloquial habits, making the language more vivid and engaging for young readers. These patterns reflect a form of syntactic normalization in translated texts, aligning with child-centered language practices. Survey results show high preference among 7–12-year-old readers for these features (more than 90% for “咚咚咚”, 83% for “抽抽搭搭”, 81% for “呢”, 74% for “把”, 76% for “得”). Chi-square tests indicated that these differences were statistically significant ( $p < 0.05$ ). Specifically, sentences incorporating reduplications (e.g., “抽抽搭搭” and “咚咚咚”) were consistently rated as more vivid and engaging due to their rhythmic and sensory qualities. Similarly, sentences using “得” constructions, such as “睡得正香”, were deemed more natural and easier to understand compared to their literal counterparts. The use of “把” constructions was also favored for its emphasis and emotional resonance, enhancing the expressive qualities of the narrative. These findings demonstrate that while both

versions adhere to Chinese grammatical norms, the translator's choices are more effective in creating a natural, lively, and accessible text for young readers.

These findings suggest that the changes made in the normalized translation are not just due to linguistic differences between English and Chinese, but are also driven by considerations of how best to cater to the reading habits and expectations of children. The simplification of sentence structure and vocabulary in the normalized translation is therefore not a mere simplification, but rather a strategic adaptation to enhance readability, comprehension, and emotional engagement for young readers.

**Translator creativity.** Translator creativity refers to the personalized interpretation and representation conducted by the translator in dialog with the original work, the original author, and the readers of the translated text. Influenced by social-historical contexts, personal perspectives, experiences, translation purposes, and other factors, translator creativity in children's literature translation involves more than just linguistic choices. It encompasses how a translator brings their background, cultural understanding, and individual experiences into the translation process. This section explores how these factors influence the creation of a translation, with particular attention to the translator's social, historical, and personal contexts, and how such creativity shapes the normalization of the translated texts.

The translator's personal experiences and background can deeply influence how they approach the translation of children's literature. For instance, the translation motivations of Chen Liating and Liu Wenlan, particularly in their work on *The Lion, the Witch and the Wardrobe*, illustrate how personal reading experiences and an admiration for classic translations can shape the translation process. Their interest in translating this work arose from their personal enjoyment of the book and a strong desire to introduce foreign literature to Chinese readers. Their guiding principle was to make the text accessible to Chinese readers while retaining the essence of the original work<sup>4</sup>. Similarly, other translators like Ren Rongrong and Yang Jingyuan bring their unique backgrounds and experiences into their work. Ren, who was significantly involved in language reform in China during the 1940s and 1950s, brought his academic background in Chinese literature and phonetic writing to inform his translations. His experience in reforming Chinese language practices allowed him to balance linguistic creativity with a desire to make the text accessible to young Chinese readers. His translations, including works like *Charlotte's Web*, are known for their careful attention to rhythm and sound, incorporating Chinese forms to reproduce the original text's cadence and humor. The use of reduplicated expressions, such as “叽嘎叽嘎” (jigajiga) to represent the sound of the mouse grinding its teeth, exemplifies Ren's playful and imaginative style. This choice of expression is not just a linguistic decision; it is a reflection of Ren's understanding of how children engage with language—through sound, rhythm, and vivid images. Yang Jingyuan draws on her own childhood experiences of imagination and creative play to inform her translations. Yang (2008, p. 11–16) reflects on her childhood experiences of seeking joy on her own while her parents were busy and unable to attend to her. Her experiences include engaging in conversations with dolls, talking to herself and creating stories, and embarking on jungle adventures, showcasing a rich imagination. Her experience of engaging in fantasy play, storytelling, and imaginary adventures as a child translated into her later work as a translator, particularly in her translation of *Peter Pan*. Her translation is known for capturing the childlike innocence and humor of the original text, using colloquial expressions and modal particles to convey the reluctance and playfulness of the child protagonist. For example, the line “I won't, I won't. Nana, it isn't six

*o'clock yet.*" is translated as "我不嘛, 我不嘛。娜娜, 还不到六点呐。 (I notma, I notma. Nana, still not yet to six o'clockna.)" This is a statement made by the young protagonist Michael when resisting going to bed. The translation adds modal particles "嘛" and "呐," expressing the child's genuine reluctance, playfulness, and exclamation. This use of familiar and relatable language serves to make the translation more accessible to young readers, and it reflects Yang's deep empathy with the child reader's perspective.

A translator's creativity is not only shaped by personal experience but also by broader social and historical contexts. Since the early 21st century, China's focus on educational reform and the standardization of language has significantly influenced the way children's literature is translated. This influence is grounded in a series of regulations and policies that emphasize linguistic standardization and alignment with national educational frameworks. For instance, the *Law on the Standard Spoken and Written Chinese Language of the People's Republic of China* (2001)<sup>5</sup>, which promotes the use of standardized Mandarin and Chinese characters, and the *Outline of China's National plan for medium and long-term Language and Writing reform and development 2012–2020*<sup>6</sup>, which emphasizes the use of standard language in educational materials, have had a profound impact on the translation of children's literature. The features of normalization in the Chinese translations of children's literature found in the previous section, such as shorter average sentence length, a higher frequency of high-frequency words, improved text readability, and more reduplicated words etc. reflect the influence of this context, with an emphasis on normalization and accessibility for young readers. For example, in the translation of *Harry Potter and the Sorcerer's Stone* (Rowling, 2019), Cao Suling and Ma Ainong demonstrated how creativity is applied in balancing linguistic and cultural adaptation to Chinese readers in this context. In an interview, Ma (2020) mentioned that "the translation of children's literature should use authentic Chinese expressions. When I translate, I make sure to pay attention to this aspect. After finishing the translation, I set the original text aside and read the translation twice purely in Chinese, from the perspective of a reader who doesn't understand foreign languages, to see if it aligns with Chinese reading habits and is at a level that children can easily understand"<sup>7</sup>. The success of these translations is reflected in their overwhelmingly positive reception on major platforms. For example, on Dangdang, Cao Suling and Ma Ainong's 2019 translation of *Harry Potter and the Sorcerer's Stone* (Rowling, 2019) has received over 200,000 reviews and maintains an impressive 99.96% positive rating. Readers frequently praise its fluency, emotional resonance, and faithfulness to the original. This demonstrates the importance of adapting to the target culture while preserving the text's original essence.

Publishers also play a crucial role in standardizing the language of children's literature translations, ensuring that translations align with national regulations and cultural norms. According to the *Regulations on Publication Administration* (2002)<sup>8</sup>, translations must adhere to established conventions for foreign names, place names, and technical terms, conforming to national language laws and modern Chinese language norms. Publishers are required to ensure clarity, accuracy, and linguistic consistency, avoiding non-standard expressions or excessive use of foreign words. Additionally, content review by publishing authorities helps guarantee that translations meet legal and quality standards. These efforts are reflected in the inclusion of works like *The Lion, the Witch, and the Wardrobe* and *Peter Pan* in official educational book lists, such as those recommended by the Ministry of Education, demonstrating how publishers align translations with both educational and regulatory requirements while maintaining high linguistic standards. At the same time, publishers respect and support the creativity of translators, enabling them to bring their unique voices to the translation process. In the case of *Harry Potter and the*

*Philosopher's Stone*, this balance is particularly evident. Based on an interview with Ma (2020), the first eight chapters were initially translated by the experienced translator Cao Suling, who later decided to discontinue due to personal convictions and age-related considerations. The publisher respected her choice and invited Ma Ainong, a translator passionate about children's literature, to complete the translation. This flexibility enabled Ma Ainong to infuse the work with her creative style, characterized by vivid colloquial expressions (e.g., "哎呀呀", "呀", "睡得正香呢"), rhythmic structures (e.g., "滴溜溜", "爬来爬去, 溜溜达达"), and emotionally resonant dialog (e.g., "可把他吓得不轻") that appeal to young readers. By accommodating individual translators' strengths and respecting their personal preferences, publishers create space for creativity within the framework of market demands and regulatory standards.

Indeed, it can be observed that the translator's creativity in children's literature is not an isolated act but is influenced by a combination of personal experiences, historical context, government regulations, and publisher guidelines. The translator adeptly utilizes characteristic Chinese structures to recreate the essence of the original work, ensuring that both language normalization and the educational objectives of the translation are met.

## Conclusion

This study, based on a corpus of classic children's literature translated into Chinese, employs primarily intra-lingual comparisons, with considerations of interlingual comparisons, to analyze the normalization features in the language of translated children's literature. The findings reveal a clear tendency toward normalization: at the overall level, the translated Chinese versions tend to feature shorter sentences, a greater repetition of diverse high-frequency words, a lower frequency of hapax legomena, and a higher textual readability. At the specific level, the frequency of four typical Chinese structures - reduplication, modal particles, "把" (BA), and "得" (DE) constructions—is either comparable to or significantly higher than that in the Chinese original corpus. The normalization observed in translated Chinese children's literature reflects a multifaceted creative process in which translators align with readers' expectations and accommodate the psychological characteristics of young readers, while also drawing on their personal experiences, responding to social and historical contexts, and adapting to the requirements and preferences of publishers. This research contributes to a deeper understanding of the linguistic features of translated children's literature, provides insights into translation universals, and highlights the complex interplay between normalization and translator creativity in crafting translations that align with educational, cultural, and market demands.

## Data availability

The data were obtained from a self-built corpus, with collection methods detailed in the Methods section. Processed datasets are available upon reasonable request to the author.

Received: 16 July 2024; Accepted: 18 June 2025;

Published online: 08 July 2025

## Notes

1 [http://www.moe.gov.cn/jyb\\_xwfb/gzdt/s5987/202004/W020200422556593462993.pdf](http://www.moe.gov.cn/jyb_xwfb/gzdt/s5987/202004/W020200422556593462993.pdf)

2 <http://bang.dangdang.com/books/bestsellers/01.41.27.00.00.00-year-2022-0-1-1>

3 <https://www.dangdang.com/> (2024-11-28) "The Lion, the Witch and the Wardrobe" (over 50,000 reviews, 99.9% positive rating), "Peter Pan" (over 20,000 reviews, 100% positive rating), "Charlotte's Web" (nearly 2.25 million reviews, 99.6% positive rating), "The Witches" (nearly 220,000 reviews, 99.9% positive rating), "Harry Potter and the

Philosopher's Stone" (over 200,000 reviews, 100% positive rating), and "The Wind in the Willows" (over 10,000 reviews, 99.8% positive rating).

4 [https://m.thepaper.cn/kuaibao\\_detail.jsp?contid=1677943&from=kuaibao](https://m.thepaper.cn/kuaibao_detail.jsp?contid=1677943&from=kuaibao)

5 [https://www.gov.cn/ziliao/flfg/2005-08/31/content\\_27920.htm](https://www.gov.cn/ziliao/flfg/2005-08/31/content_27920.htm)

6 [http://www.moe.gov.cn/srcsite/A18/s3127/s7072/201212/t20121210\\_146511.html](http://www.moe.gov.cn/srcsite/A18/s3127/s7072/201212/t20121210_146511.html)

7 [https://www.douban.com/note/780957885/?\\_i=5393803br19-cW,4106509br19-cW](https://www.douban.com/note/780957885/?_i=5393803br19-cW,4106509br19-cW)

8 [https://www.gov.cn/gongbao/content/2002/content\\_61879.htm](https://www.gov.cn/gongbao/content/2002/content_61879.htm)

9 [https://www.gov.cn/zhengce/2016-10/12/content\\_5713806.htm](https://www.gov.cn/zhengce/2016-10/12/content_5713806.htm)

## References

Anthony L (2020) AntConc (Version 3.5.9). Waseda University, Tokyo

Baker M (1993) Corpus linguistics and translation studies: implications and applications, in Text and technology: In honor of John Sinclair, eds. M Baker, G Francis and E Tognini-Bonelli, Benjamins: Amsterdam/Philadelphia, 233–250

Baker M (1996) Corpus-based Translation Studies: the Challenges that Lie Ahead. Terminology, LSP and Translation: Studies in Language Engineering in Honour of Juan C. Sager, eds H Somers, John Benjamins, 175–186

Blum-Kulka S (1986) Shifts of Cohesion and Coherence in Translation. Interlingual and Intercultural Communication: Discourse and Cognition In Translation and Second Language Acquisition Studies, eds J House, S Blum-Kulka, Gunter Narr, 17–35

Blythe H, Liversedge S, Joseph H, White S, Rayner K (2009) Visual information capture during fixations in reading for children and adults. *Vis Res* 49(12):1583–1591

Cao WX (2014) Cao Wenxuan on Children's Literature. Dolphin Books, Beijing

Chesterman A (2010) Why study translation universals? *Acta Transl Helsing* 1:38–48

Chesterman A (2014) Translation studies forum: universalism in translation studies. *Transl Stud* 7:82–90. <https://doi.org/10.1080/14781700.2013.828904>

Chesterman A (1997) Memes of Translation: The Spread of Ideas in Translation Theory. John Benjamins

De Sutter G, Delaere I, and Plevaerts, K (2012) Lexical lectometry in corpus-based translation studies. Oakes, M & Ji, M Quantitative methods in corpus-based translation studies: A practical guide to descriptive translation research 51, 325–345

Du S, Wang X, Ding C (2021) A Corpus-based Study of Normalization in English Version of Jade King. *J Univ Shanghai Sci Technol* 3:242–248

Hu J, Guo H, Qi Y (2021) A Phraseological Examination of Normalization of Translation Universals. *Chin Transl J* 4:141–149

Hu K (2009) Corpus-based Study of BA-construction in the Chinese Versions of Hamlet by Shakespeare. *Foreign Lang Res* 1:111–115

Hu X (2010) A corpus-based multi-dimensional analysis of the stylistic features of translated Chinese. *Foreign Lang Teach Res* 6:451–458

Hu X, Zeng J (2011) Hybridization of Translated Chinese as Observed in the Use of "ba" (把) Constructions. *Foreign Lang Res* 6:69–75

Hu X (2006) A Corpus-based Study on the Translational Norms of Contemporary Chinese Translated Fiction. Diss. East China Normal University

Huang Y (1993) An exploration of the art of duplicated words in Li Qingzhao's "Slow Slow Tune". *Contemp Rhetor* 3:26–27

Jia J, Afzaal M, Naqvi S (2022) Myth or reality? Some directions on translation universals in recent corpus based case studies. *Front Psychol* 13:902400. <https://doi.org/10.3389/fpsyg.2022.902400>

Jiang F (2013) New Children's Literature Course. Zhejiang University Press, Hangzhou

Joseph H, Liversedge S, Blythe H, White S, Rayner K (2009) Word length and landing position effects during reading in children and adults. *Vis Res* 49(16):2078–2086

Ke F (2003) The characteristics and English translation of Chinese Ba-construction. *Foreign Lang Their Teach* 12:1–5

Kenny D (2001) Lexis and Creativity in Translation. St. Jerome Publishing, Manchester

Laviosa S (1998) The English comparable corpus: A resource and a methodology, in Unity in diversity: Current trends in translation studies. eds. L Bowker, M Cronin, D Kenny and J Pearson (St. Jerome: Manchester), 101–112

Lei L, Yaoyu W, Kanglong L (2024) AlphaReadabilityChinese: A tool for the measurement of readability in Chinese texts and its applications. *Foreign Lang Their Teach* 46(1):83–93

Li C (1986) An Introduction to the Psychological Development of Children from Birth to Adolescence. Southwest China Normal University Press, Chongqing

Li J (2016) Passion sprouts wisdom: Selected papers on contextual education by Li Jinlin. Educational Science Publishing House, Beijing

Li X (2021) The effect of word length in Chinese children's natural sentence reading and its development. Thesis. Tianjin Normal University

Ma A (2020) The translator of Harry Potter, Ma Ainong—Translating Children's Literature with Authentic Chinese. [https://www.douban.com/note/780957885/?\\_i=5393803br19-cW](https://www.douban.com/note/780957885/?_i=5393803br19-cW)

May R (1997) Sensible elocation: How translation works in and upon punctuation. *Translator* 1:1–20. <https://doi.org/10.1080/13556509.1997.10798985>

Pan F (2014) A Corpus-based Study of Normalization in Chinese-English Conference Interpreting. *Foreign Lang Lit* 6:144–150

Puurinten T (2003) Genre-specific Features of Translationese? Linguistic differences between translated and non-translated finnish children's literature. *Lit Linguist Comput* 4:389–406. <https://doi.org/10.1093/linc/18.4.389>

Pym A (2008) On Toury's laws of how translators translate. In: Pym, A, Shlesinger, M, and Simeoni, D *Beyond descriptive translation studies. Investigations in homage to Gidon Toury*. Amsterdam: Benjamins. 311–328

Rowling J (2019) Harry Potter and the Philosopher's Stone. Translated by Cao Suling and Ma Ainong. People's Literature Publishing House, Beijing

Santos DD (2024) Normalisation in self-translation: a corpus-based analysis of An Invincible Memory by Joao Ubaldo Ribeiro and a review of the #LancsBox concordance. *Cad De Traducao* 1:1–18. <https://doi.org/10.5007/2175-7968.2024.e98752>

Tully EJ (2014) Translation universals and polygenesis: implications for textual criticism. *Bible Translator* 65:292–307. <https://doi.org/10.1177/2051677014553534>

Vanderauwera R (1985) Dutch Novels Translated into English: The Transformation of a Minority Literature. Radopi

Wan Y (2019) Commentary on translation and publishing of children's books since China's reform and opening-up. *Res Hist Publ China* 3:32–47

Wang K, Hu X (2008) A parallel corpus-based study on lexical features of translated Chinese. *Chin Translators J* 6:16–21

Wang A (2016) On the language identifiability of translated children's literature: a research into the target readers' response. *J Anqing Normal University (Soc Sci Ed)* 2:85–89

Wu J (2019) On Normalization in Chinese-English Consecutive Interpreting - an Empirical Study Based on Novice Interpreters and Professional Interpreters. Thesis. Central China Normal University

Wu K, Li D (2021) Normalization, Motivation, and Reception: A Corpus-Based Lexical Study of the Four English Translations of Louis Cha's Martial Arts Fiction. *New Perspectives on Corpus Translation Studies*, edited by VX Wang, L Lim and D Li, Springer, 2021, 181–199

Yang J (2008) A Different Kind of "Co-translation". A Book and a World 2, edited by Xianyi Y Beijing: Kunlun Press, 11–16

Yu Y (2014) A Corpus-based Research of Language Use in 3–6 years-old Translated Chinese Children's Picture Books. Thesis. Changchun: Jilin University

Zang G (2010) A Corpus-based Study of Language Use in Translated Chinese Children's Literature. Thesis. Qufu Normal University

Zhang H, Liu Q (2002) ICTCLAS. Institute of Computing Technology, Chinese Academy of Sciences, Beijing

Zhang S (2010) Drama in Education. China Drama Press, Beijing

Zhang Y (2000) Relation Between de-Sentence and Writing Style. *J Henan Univ (Soc Sci)* 1:105–108

## Author contributions

The author confirms being the sole contributor of this work and has approved it for publication.

## Competing interests

The author declares no competing interests.

## Ethical approval

Ethical approval was not required for this study in accordance with Article 39 of China's *Ethical Review Measures for Biomedical Research Involving Human Subjects* (Ministry of Health, 2016)<sup>9</sup>, which exempts research involving minimal-risk anonymous surveys that collect only non-identifiable, aggregated data. This study employed fully anonymized questionnaires to gather aggregated preference data on literary translations, with no collection of personally identifiable information or individual-level data analysis.

## Informed consent

All participants were informed through written instructions at the beginning of the questionnaire (conducted from December 1 to 30, 2024) about the anonymous and voluntary nature of this study. This survey component was added during the revision process to address specific reviewer suggestions, complementing the original corpus-based analysis. By proceeding to complete and submit the survey, participants confirmed their consent to participate.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1057/s41599-025-05379-6>.

**Correspondence** and requests for materials should be addressed to Yang Han.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025