Journal Of the Iraqia University (76-4) 31 July (2026)

ISSN(Print): 1813-4521 OnLine ISSN:2663-7502

Journal Of the Iraqia University

available online at

https://iasj.rdd.edu.iq/journals/journal/view/247

The Effect of Audience-Oriented Prompting on AI Translation of Children's Literature: A Comparative Study of Three Stories from Kalila wa Dimna

Fatima Awad Mikhlif

Affiliation: University of Tikrit, College of Arts, Department of English

Email: fatima.mukhlif122@tu.edu.iq

Abstract

Recent advances in Artificial Intelligence (AI) have significantly increased the quality of machine translation. But there are questions about the ability of AI systems to tailor translations to the needs of particular audiences, especially children. The present study investigates the effect of audience-oriented prompting on AI-based literary translation by comparing the prompted and non-prompted English translations of three short stories selected from Kalila wa Dimna: Bab al-Qird wa al-Ghaylam, Bab al-Labwah wa al-Shihr, and Bab al-Nasik wa Ibn ‘Irs. The study is a comparative qualitative analysis which evaluates the translations based on two variables: translation accuracy at the whole-text level and stylistic appropriateness for children aged 8–12. The findings show that the main events, character relations, narrative structure and moral lessons of the source texts were mostly maintained in the translations, both prompted and non-prompted. However, the prompted translations were consistently more stylistically appropriate for children . Both translation conditions achieved similar levels of whole text accuracy. The increase in overall accuracy was rather modest, but the increase in stylistic appropriateness for children was considerable. The prompted translations used simpler vocabulary, easier sentence structure and a more accessible narrative tone, whereas the non-prompted translations were more literal, more formal and less audience-oriented. The research indicates that audience-based prompts can enhance the communicative effectiveness of AI-generated literary translation by increasing their readability and appropriateness to the target audience without losing the central meaning of the source text. Our findings emphasize the importance of prompt engineering in AI-assisted translation, particularly when translating for a specific audience.

Keywords: Artificial Intelligence Translation; Prompt-Based Translation; Prompt Engineering; Children's Literature Translation; Kalila wa Dimna; Translation Quality Assessment; Audience-Oriented Translation

1. Introduction

Artificial Intelligence (AI) is gaining more and more importance in the field of translation, especially in the last years with the appearance of Neural Machine Translation (NMT) systems. The systems can generatetranslations that are fluent and mostly correct for a lot of languages and kinds of text. Therefore, AI-based translation is now widely used both in professional and everyday life. But there are still significant challenges, especially in cases where an awareness of audience, tone, and stylistic variation are needed for translation. One area in which these challenges are especially visible is the translation of literary texts and particularly children’s literature. Children’s translation is not only a matter of expressing the basic meaning of the source text. This also means adapting language to the cognitive and language development of young readers, often with simpler vocabulary, more direct sentence construction and a livelier narrative voice. The right translation may not be stylistically appropriate and therefore may not be communicative with its audience. Prompting is a new way to communicate with AI. Prompting is simply when users provide direct instructions to change the output of AI systems. Promptings can be used in translation tasks to indicate the purpose of the translation, the target audience, and stylistic features like tone and readability. This leads to an important question: if prompt-based translation can improve not only the accuracy of AI output, but also its appropriateness for particular audiences like children. The role of prompting in literary translation, particularly in terms of audience-oriented adaptation, has received less attention in previous research, whereas the accuracy and performance of AI translation systems have been the focus of earlier research. Most of the related work

focuses on the accuracy of sentence translations or the overall quality of machine translation, with no systematic comparison of prompted versus non-prompted outputs regarding both semantic and stylistic variables. This study seeks to bridge this gap by exploring the impact of prompting for AI-based literary translation for children. It does this by means of a comparative analysis of prompted and non-prompted English translations of three short stories from Kalila wa Dimna. The analysis is based on two major variables, the overall accuracy of translation and stylistic suitability for children 8-12 years. The importance of this study lies in its attempt to link the AI translation with the functional and audience-based approaches in translation studies. The paper explores the impact of prompting on meaning transfer and stylistic adaptation, and enhances our knowledge of how to better utilize AI in literary translation, especially when the primacy of the target audience is acknowledged.

2. Theoretical Framework

2.1 Machine Translation and the Development of AI-Based Translation

Machine Translation (MT) has changed from rule-based systems to statistical systems and now Neural Machine Translation (NMT) over the past decades. Early RBMT systems were based on linguistic rules and bilingual dictionaries, while SMT systems were based on probabilistic models trained on large bilingual corpora. However, both approaches were lacking context, fluency and stylistic variation (Koehn, 2020). Neural Machine Translation has been a significant change in the field. NMT systems utilize deep learning models to translate entire sentences instead of fragments, which leads to more fluent and contextually relevant translations. NMT can better model the semantic relations and generate more natural target texts compared with previous systems. However, these advances do not prevent NMT systems from still presenting problems of stylistic adaptation, audience awareness and coherence at the discourse level (Rivera-Trigueros, 2022). These problems are all the more acute in the case of literary translation where translation is not merely a matter of transferring the meaning but one of tone, style and cultural adaptation. This makes AI-based translation especially vulnerable to external manipulation, such as prompting.

2.1.1 Large Language Models and Translation

Recent advances in Artificial Intelligence have led to the replacement of traditional Neural Machine Translation (NMT) systems with Large Language Models (LLMs) such as ChatGPT. Whereas traditional NMT systems are designed for translation, LLMs are general purpose language models that can translate using natural-language instructions and promptsAtaman (2025) argues that LLMs mark a new phase in the development of machine translation, since they can integrate translation with more general language-generation capabilities. Unlike translation-oriented architectures, LLMs can adapt their output to user instructions, communicative goals, stylistic preferences and audience needs. This flexibility is particularly relevant for literary translation and other audience-oriented translation tasks. Traditional MT systems are usually focused on semantic transfer, while LLMs can also follow instructions on tone, readability, and style adaptation. Therefore, prompt design is an important factor affecting the quality of the translation and its communicative effectiveness. The current study thus explores AI translation in the context of LLM-based translation where prompting is a means of guiding the translation towards the needs of a particular readership.

Section 2.2 Prompt-Based Translation and Audience Adaptation

Recent progress in Large Language Models (LLMs) has demonstrated the importance of prompting as a way of guiding AI outputs. For translation tasks, prompts are a way for users to provide instructions on purpose, tone, readability level, and intended audience. The AI system can be directed towards a translation fulfilling specific communicative needs rather than generating a generic translation. Studies on prompt engineering have shown that LLMs are able to change their vocabulary, sentence structure, and discourse style to suit the user's query. Brown et al. (2020) showed the general ability of large language models to respond by shaping the features of the generated translation while preserving the core meaning of the source text. Yamada (2023) further argues that prompts may have a similar role to that of translation instructions provided to human translators in the past. By defining the communicative goal of the translation, prompts prompt the model to change lexical choice, readability and discourse style in a way that meets audience expectations. This observation is particularly relevant for the present study, which investigates the impact of audience-oriented prompting on translations for children. Similar findings were reported by He (2024) when examining the impact of translation briefs and translator-oriented personas on ChatGPT-generated translations. The research showed that the stylistic and functional features of the translated text can be affected by the prompts with the information about communicative purpose, target readership and translation requirements. These results suggest that prompt

design could be a practical means of integrating translation-oriented instructions into AI-assisted translation workflows. Sharkas (2025) also showed that the presence of translation brief elements in prompts can have a strong impact on the linguistic and stylistic features of LLM-generated translations. Target audience, communicative purpose, and readability requirements were among the elements that were found to direct the model to outputs that better match the intended function of the translation. This supports the view that prompting can act as a kind of translation assistance similar to the instructions that are normally given to human translators. Pourkamali et al. (2024) further provided evidence by studying the effect of different prompting strategies on LLM-based machine translation. Their findings suggest that prompt configuration can affect translation quality, lexical choice and stylistic realization across language pairs. The work highlights the importance of prompt design as an influencing factor of the translation output, not only for the initiation of the translation task. Audience-oriented prompting is especially helpful in the case of literary translation, as different readerships have different linguistic and stylistic features. Understanding who you are writing for can help the AI produce better results that are more useful and relevant to your readers. The adaptation of the audience is of particular importance in children's literature, as young readers differ from adults in linguistic competence, cognitive development and reading experience. This suggests that prompts referring to child readers may lead to a simpler vocabulary, more transparent syntax, and a more accessible narrative style. This way prompting provides a practical way of tailoring AI-generated translations to specific audiences.

2.3 Translation of Children's Literature

The translator of children's literature has special problems, for he must not only translate faithfully but adapt to the age of his readers. The translator has to take into account the simplicity of the vocabulary, the structure of the sentences, the tone of the narrative and the clarity of emotion. For example, Oittinen (2000) states that translating for children is basically different from translating for adults, since the text has to be adapted to the cognitive and linguistic capacity of children. This normally means reduction, explication and stylistic adaptation. Lathey (2016) emphasises the need to find a balance between fidelity to the source and readability and attractiveness, much as in the translation of children's literature. A translation that is too literal may retain meaning but fail to speak to child readers. Recent scholarship on successful translation for children continues to emphasize the importance of audience adaptation. Contemporary research sees readability, accessibility and age appropriateness as important evaluation criteria, because young readers process narrative texts differently to adults. Thus, translations often use a less difficult vocabulary, less complex structures and a clearer story line, but still preserving the essential content and educational value of the source text. These principles are especially relevant for AI-assisted translation. When AI systems are to translate for child readers, the evaluation criteria are semantic accuracy, readability, accessibility and adaptation to the audience. This view forms the basis of the present study which is concerned with the preservation of meaning and the stylistic suitability for children. These principles apply directly to AI translation. Even without explicit instruction, AI systems can produce linguistically correct translations that are stylistically inappropriate for children. This suggests the prompting as a process of audience adaptation.

2.4 Translation Quality Assessment

The quality assessment of translation (TQA) is the important problem of the translation studies. The traditional models measure translation in terms of accuracy, equivalence and error analysis. However, recent approaches are more and more concerned with functional adequacy. Functional adequacy is the extent to which a translation is able to serve its purpose (House, 2015). Recent research on LLM prompting has empirically validated the link between audience specification and translation performance, illustrating how the components of the translation brief contained within prompts can affect the extent of audience adaptation obtained in the ultimate translation (Sharkas, 2025).

The study uses two variables in the context of this study:

1. Accuracy of translation as a whole

This is the degree of preservation of the translation of:

- main events
- character relationships
- narrative structure
- moral meaning

This is consistent with the notion of global semantic accuracy which is concerned not with isolated lexical errors but with the whole message.

2. Stylistic suitability for children

That means:

- simplicity of vocabulary
- sentence clarity;
- tone of narration
- Emotional availability

This variable is an indicator of functional adequacy because it assesses the appropriateness of the translation for the target audience. Recent research suggests the need to go beyond sentence accuracy and towards discourse and audience-based evaluation in the context of AI translation (Zappatore et al. 2024).

2.5 Functionalist Approach and Audience-Oriented Translation

Functionalist theories like Skopos Theory maintain that the translation should be determined by the purpose (skopos) and the target audience. Vermeer (1989) states that a translation is successful if it works in the target situation.

The target audience in this study is children 8-12 years old. Therefore, the translation should:

- be comprehensible
- be interesting
- meet the reader's level of cognition

The translation based on the prompt is direct evidence to the approach that the skopos of the translation is explicitly defined. The prompt functions as a working guide for the translation output by guiding the AI to produce a child-friendly version.

This is why prompting was more effective than basic accuracy to support stylistic appropriateness in the present study.

2.6 AI Translation in Literary Contexts

AI translation preserves the general meaning fairly well, especially in narrative texts. Yet literary translation requires additional layers of interpretation , including tone , style and rhetorical effect . Recent works indicate that LLM-based systems may provide greater flexibility than traditional NMT systems in literary contexts, as they can customize their output based on explicit user instructions and communicative goals (Ataman, 2025).

Research shows that AI systems will likely produce:

- literal and accurate translations
- grammatically correct but stylistically neutral output;

Without guidance, AI will tend to default to a formal, generic style, which may not be appropriate for specific audiences (Koehn, 2020; Rivera-Trigueros, 2022).

This is consistent with findings of the present study:

- Non-prompted translation = accurate but formal
- Prompted translation = accurate + adapted style

Thus, prompting plays a key role in bridging the gap between semantic accuracy and communicative effectiveness.

3. Methodology

3.1 Research Problem

With the rapid development of AI-based translation, especially Neural Machine Translation, these systems have become able to produce translations that are generally accurate in terms of meaning. But there is a big problem that remains unsolved: AI translation is often insensitive to the target audience, especially in literary contexts such as children's literature. Children's literature must be semantically accurate but also stylistically adapted, e.g. simple vocabulary, readable sentence structure, engaging narrative tone The unsupervised use of AI systems often produces translations that are accurate but formal and less suitable for young readers.

This raises an important problem:

To what extent can prompting improve AI translation, especially in terms of adapting the text to a specific audience such as children?

3.2 Research Objectives

1. This study aims at:

2. study the effect of prompting on AI-based translation

3. compare prompted and unprompted translations of literary texts

4. assess whether prompting improves:

- overall translation accuracy

-stylistic appropriateness for children

5. decide which variable is more affected by prompting

6. assess the function of prompting as a device for audience-oriented translation

3.3 Research Questions

The study will attempt to answer the following questions:

1- Does prompting improve the accuracy of AI translation at the holistic level

2- Does prompting improve stylistic suitability for children

3- What's the difference between prompted and non-prompted translations

4- What is more influenced by prompting – accuracy or style?

3.4 Research Hypotheses

The study proposes the following hypotheses based on the theoretical framework:

1. Whole-text accuracy will be similar for prompted and non-prompted translations.

2. Prompts for translations will be more stylistically appropriate for kids.

3. Stylistic adaptation will be more affected by prompting than overall accuracy.

3.5 Research Design

This study adopts a **comparative qualitative analytical design** supported by structured evaluation.

The research compares two translation conditions:

- **prompted translation** (AI guided by a child-oriented prompt)
- **non-prompted translation** (AI used without any instruction)

Each condition is applied to the same set of texts, and the outputs are analyzed using identical evaluation criteria.

3.6 Data Selection

The data of the study consist of **three short stories from Kalila wa Dimna**:

1.Bab al-Qird wa al-Ghaylam

2.Bab al-Labwah wa al-Shihr

3.Bab al-Nasik wa Ibn 'Irs

These stories were selected because:

- they are **narrative and moral in structure**
- they contain **dialogue and embedded stories**
- they are suitable for **children's adaptation**
- they represent classical Arabic literary style

3.7 Data Collection Procedure

Each story was processed in two conditions:

1. Prompt Translation

The prompt design is inspired by recent research that underscores the importance of translation briefs and audience-sensitive instructions in influencing the translation output of LLMs (He, 2024).

The AI was provided a specific prompt to :

•translate for kids 8-12 years old

•utilise simple and clear vocabulary

•keep the tone of natural storytelling

•maintain the overall meaning and events

2. Translation without hints

The same text translated:

•without direction

•using the default AI behaviour

Both times, the input texts and the AI system were identical, for consistency.

This control over the translation conditions is in line with recent studies on the impact of prompt configuration on LLM translation performance (Pourkamali et al., 2024).

3.8 Variables of the Study

This study deals with two main variables:

1. Accuracy of Whole Translation

This is related to the degree of fidelity of the translation:

- significant events
- relationships with others
- narrative structure
- Moral of the story

This variable measures global meaning, not individual word errors.

2. Stylistic Suitability for Children

This refers to the degree to which the translation:

- Uses basic vocabulary
- Uses clear and readable sentences
- Uses an engaging narrative style
- Presents emotionally accessible material

This variable is the measure of audience adaptation.

3.9 Evaluation and Scoring Method

The evaluation was done through a structured qualitative evaluation on two well-defined variables: accuracy in the whole text translation and stylistic suitability for children aged 8–12. The study is exploratory and based on a limited corpus of three literary stories, so the scoring is used as an analytical support tool, not as a claim of statistical generalizability. Each translation was scored by direct comparison with the Arabic source text when it was produced under the same condition, and with the translation produced under the other condition. The evaluation was based on whether the translation kept the main narrative content and presented it in a form suitable for child readers.

A five-point scale was used:

5 = very high quality
4 = high quality
3 = moderate quality
2 = low quality
1 = very low quality

To minimize subjective judgement, the scores were assigned according to the rubrics shown below. The same criteria were applied to all three stories and both prompted and non-prompted translations. A limitation of the study is that it was carried out by a single evaluator and thus the findings should be interpreted as indicative and exploratory. Future studies may increase the reliability of the assessment by using multiple raters and calculating inter-rater agreement.

The researcher evaluated all translations with the presented rubric in Section 3.9.1. The same evaluation procedure was applied to all stories and translation conditions.

3.9.1 Evaluation Rubric

To minimize subjective judgement, the scores were assigned according to the rubrics shown below.

A. Accuracy of Translation as a Whole

This variable assesses whether the translation retains the meaning and narrative structure of the source text. Four indicators are considered:

1.Preservation of main events
2.Keeping the relationship between characters
3.Maintaining coherence and the sequence of the narrative
4.Maintaining the moral and the overall communicative content The following scale was used for scoring:

Score	Description
5=	All four indicators are preserved completely and clearly
4=	Minor loss of detail, but overall meaning remains intact
3=	General meaning preserved, but one or more important elements weakened

Score Description

2= Significant loss of narrative structure or moral message

1= Core meaning not preserved

B. Style Appropriate for Children This variable measures the extent to which the translation is appropriate for readers aged between 8 and 12 years. Four indicators were considered:
1. Simplicity of vocabulary
2. Readability of sentences
3. Flow of story and narrative engagement
4.Emotional clarity and availability The score was assessed on the following scale:

Score Description

5= Highly accessible and consistently suitable for children aged 8–12

4= Generally suitable with minor stylistic difficulty

3= Moderately suitable, but contains noticeable formal or abstract language

2= Difficult for children because of vocabulary or sentence complexity

1= Largely unsuitable for child readers

Scoring Procedure

For each story the evaluator first viewed the translation according to the four indicators of accuracy and then according to the four indicators of stylistic suitability. The final score was based on the translation's performance as a whole in relation to the indicators, not on the presence or absence of isolated lexical differences.

We applied this procedure to all prompted and non-prompted translations in the same way.

3.10 Data Analysis Method

The study uses **comparative textual analysis**, which includes:

1.comparing prompted vs non-prompted translations
2.identifying differences in:
- clarity
- wording
- tone
- structure

3.evaluating both versions based on the two variables
4.summarizing results in structured tables
5.The analysis is not at the level of isolated sentences but at the level of whole text performance. The analysis focuses on functional and communicative effectiveness at the discourse level, not on sentence-by-sentence counting of errors, in order to reflect the nature of literary translation for children.

3.11 Reliability and Validity

The evaluation was performed by one researcher to ensure consistency across all evaluations, but it is acknowledged that qualitative translation evaluation involves some subjective judgment. Future studies could involve multiple evaluators and measures of inter-rater reliability to enhance the robustness of the findings. To guarantee reliability:

- Identical texts were used in both conditions
- used identical evaluation criteria
- the same prompt was used for all stories
- To make it valid: evaluation was focused on well-defined variables
- analysis backed up by textual examples
- Uniform performance across the three stories

3.12 Limitations of the Study

The study has a number of limitations:

- small corpus (three stories only)
- one AI system applied
- assessment by a single researcher
- only consider narrative literary texts

These limitations point to the necessity of further investigation with larger data sets and multiple evaluators. The present study uses publicly available literary texts and AI-generated translations . No human subjects were involved and no personal or sensitive data were collected. Hence, the study does not entail any ethical risks. This study is exploratory and not intended to produce statistically generalizable results. What it tries to do though is to provide a first qualitative investigation into the impact of audience-oriented prompting on AI-assisted literary translation.

4. Data Presentation, Analysis and Results

4.1 Data Presentation

Due to the scale of the study dataset, the full corpus has been compiled in a supplemental document, including the source language stories as well as their prompted and non-prompted translations. To see the full dataset being used, click the link or scan the QR code provided:
https://docs.google.com/document/d/108ote9Wr7hJWqmHmzAzjGeAbSPmFC5Bb/edit?usp=sharing&ouid=103916223953976657886&rtpof=true&sd=true

4.2 Analysis of the Prompted and Non-Prompted Translations of *Bab al-Qird wa al-Ghaylam*

The narrative structure of Bāb al-Qird wa-al-Ghaylam is completely consistent in each translation model. In both versions the monkey is driven out of his tribe, forms a firm friendship with the tortoise, is threatened with having his heart cut out by the tortoise, cunningly engineers his escape by means of deceptive talk, and tells the tale of the lion, the jackal and the donkey as a warning before the philosopher ends the pedagogical example. However, the non-prompted and audience-prompted translations differ significantly in their linguistic and stylistic choices, particularly in terms of the way they approach fluency, vocabulary selection and audience appropriateness. The human evaluation scores for translation accuracy overall are shown in Table 1. The comparative metrics for stylistic suitability for the children aged 8-12 are depicted in Table 2. As for the first variable, accuracy of translation as a whole, both versions are at an identical, high level, keeping all core plot developments, character relationships and the final moral framework without any omissions or distortions. The only difference between the two outputs lies in the way that this preserved meaning is expressed and presented to the target reader. The non-prompted translation clearly shows a tendency to be faithful to the source text, in syntax and semantics, staying closer to the formal and rhetorical structure of the Arabic prose. But it produces a clumsy and literal translation, such as “seeks his need”, “attains a thing”, “acts with deliberation”, and “it is not fitting for a friend to withhold from his companion either counsel or benefit”. The choices made result in a text which is over-formal, heavy and structurally dense which decreases readability for a younger target audience. By contrast, the audience-driven translation optimizes the narrative flow for children by simplifying syntax and favoring familiar lexical alternatives, as shown in the textual evidence in Table 3. For example, the prompted version replaces abstract or archaic language with child-friendly language: the Arabic clause al-naṣīḥah wa-al-manfaʿah (النصيحة والمنفعة) is translated as “advice or help” instead of “counsel or benefit,” and “seeks his need” is translated intuitively as “what one wants”.

Table 1. Comparison of the Prompted and Non-Prompted Translations in Terms of Translation Accuracy as a Whole

Criterion	Prompted Translation	Non-Prompted Translation	Difference
Preservation of main events	Preserves all major events of the source story	Preserves all major events of the source story	No major difference
Preservation of character relations	Maintains monkey–turtle friendship, wife's demand, betrayal attempt, escape, and final moral structure clearly	Maintains the same character relations and narrative sequence	No major difference
Preservation of embedded story	Retains the lion, jackal, and donkey episode fully and coherently	Retains the embedded tale fully and accurately	No major difference
Preservation of narrative sequence	Maintains the order and progression of events throughout the story	Maintains the same sequence and progression of events	No major difference
Faithfulness to overall moral lesson	Keeps the moral example of gaining something and failing to keep it	Keeps the same moral lesson and closing wisdom	No major difference
Overall judgment	Highly accurate as a whole text	Highly accurate as a whole text	Both translations achieve a similarly high level of whole-text accuracy

Suggested score

- Prompted translation: **5/5**
- Non-prompted translation: **5/5**

Table 2. (Story 1) Comparison of the Prompted and Non-Prompted Translations in Terms of Stylistic Suitability for Children

Criterion	Prompted Translation	Non-Prompted Translation	Difference
Vocabulary simplicity	Uses simpler expressions such as "what one wants" and "advice or help"	Uses more formal phrases such as "his need" and "counsel or benefit"	Prompted version more child-friendly
Sentence readability	Sentences are smoother and easier to follow	Sentences are more formal and sometimes heavier	Prompted version more readable
Narrative tone	More natural storytelling tone for younger readers	More formal and text-like tone	Prompted version more engaging
Accessibility of dialogue	Dialogue sounds clearer and easier for children to understand	Dialogue remains understandable but more rhetorical and adult-like	Prompted version better suited to children
Emotional immediacy	Emotional reactions are easier to grasp	Emotional content is preserved but phrased more formally	Prompted version stronger
Suitability for ages 8 to 12	More appropriate for child readers	Less appropriate because of abstract and formal diction	Prompted version clearly better
Overall judgment	Highly suitable for children	Moderately suitable, but not ideal for children	Prompted version clearly superior

Suggested score

- Prompted translation: **5/5**
- Non-prompted translation: **3/5**

Table 3. (Story 1) Textual Evidence Supporting the Comparative Analysis

Source Meaning	Prompted	Non-Prompted	Observation
النصيحة والمنفعة	advice or help	counsel or benefit	Simpler vocabulary in prompted version
وصف الصديق	A true friend should never keep advice or help from his companion	It is not fitting for a friend to withhold from his companion either counsel or benefit	More natural syntax and readability
طلب الحاجة	what one wants	his need	More familiar vocabulary
الحوار السردي	smoother conversational exchanges	more formal rhetorical structure	Greater child accessibility

These examples demonstrate that the prompted translation systematically simplifies the vocabulary and the sentence structure, while maintaining the original meaning. The non-prompted version is semantically correct, but uses more formal and literary expressions. This further supports the conclusion that prompting leads to a stronger increase in audience-oriented adaptation than in semantic preservation.

4.3 Concluding comparison

In comparing the two English versions of Bab al-Qird wa al-Ghaylam, we find that both translations are accurate at the whole-text meaning level, since both have preserved the source plot, the embedded narrative and the moral lesson. But the prompted translation is better in two respects: it tells the story more clearly in English and it is more successful in adapting the style for children. The non-prompted translation is still correct but more literal, more formal, and less understandable for younger readers. So, in this case, prompting makes the translation not only more readable but also more successful in its functional purpose for its intended audience.

4.4 Analysis of the Prompted and Non-Prompted Translations of *Bab al-Labwah wa al-Shihr*

The narrative and ethical framework of Bāb al-Labwah wa-al-Shihr is successfully preserved in both conditions of translation. The comparison shows that both translations keep the general structure of the Arabic source text, that is the king's request for a moral example, the philosopher's explanation, the lioness's loss of her cubs, the jackal's moral rebuke, the lioness's change of behavior and the final wisdom stated by the philosopher to the king. In both versions the main plot is retained: the lioness loses her cubs, the jackal reminds her that she is now suffering what others once suffered because of her own violence, she realizes her guilt, gives up meat, then fruit, and finally turns to ascetic devotion. However, there are important differences in the way that each translation renders the stylistic presentation of this moral transformation. The overall translation accuracy human evaluation scores are shown in Table 4. Table 5 shows comparative metrics on stylistic suitability for children aged 8–12. Both translations work at a similar high level for the first variable – accuracy of translation as a whole. The prompted and non-prompted versions maintain the central story and ethical framework of the source text. Neither translation leaves out or distorts substantive portions that affect the overall meaning of the story. As such, both translations may be considered highly accurate at the whole-text level. The primary difference is not accuracy, but how the preserved meaning is presented to the reader.
The non-prompted translation is closer to the formal texture and the rhetorical dense style of the Arabic source. "Stops hurting others because of the hurt that comes to him," "their power over things," and "the result of every action is either reward or punishment" are all close to the original structure but sound more formal and less natural in English narrative prose for children. This literalness, however, does not detract from the accuracy of the translation as a whole but conveys the meaning in a more ponderous and text- bound way.
On the contrary, the translation prompted by the audience successfully carries out lexical and stylistic simplifications in order to reduce conceptual complexity for 8-12 year-old readers, as detailed in the textual evidence in Table 6 . For instance, the prompted version uses the phrase "harming others" while the non-prompted version states the same idea as "committing injustice and aggression". In a similar vein it makes moral ideas more direct replacing the ominous phrase 'before the evil consequence descends' with 'when the punishment finally comes'. Both translations correctly preserve the lesson of "as you judge, so shall you be judged". The prompted translation uses simpler and more readable language to express the same message. Moreover the prompted model greatly improves the flow of the narrative and emotional clarity. The non-prompted phrase in describing mother's anguish "She screamed, cried out, and rolled over on her back and

belly" sounds formal and not so direct. It does better than this to paint a more vivid picture of emotion: "She screamed and cried and rolled over and over in grief. Again, the jackal's advice is in a more natural, child-friendly rhythm ("Bear from others patiently that which others once bore from you patiently") than the stiffer option of the default translation ("Endure from others as others once endured from you"). The examples show that the translation prompted is always characterized by a moral and emotional content, expressed in a clearer and more accessible language. This indicates that prompting contributes more to audience-oriented adaptation and readability than to basic semantic preservation.**Table 4. (Story 2) Comparison of the Prompted and Non-Prompted Translations in Terms of Translation Accuracy as a Whole**

Criterion	Prompted Translation	Non-Prompted Translation	Difference
Preservation of main events	Preserves all main events of the source story	Preserves all main events of the source story	No major difference
Preservation of moral structure	Retains the movement from grief to recognition, repentance, and moral lesson	Retains the same moral structure clearly	No major difference
Preservation of character roles	Maintains the lioness as wrongdoer turned penitent and the jackal as moral commentator	Maintains the same roles and relations	No major difference
Preservation of narrative sequence	Maintains the sequence from loss to moral transformation clearly	Maintains the same narrative sequence	No major difference
Faithfulness to final ethical lesson	Keeps the message of reciprocal justice and fairness to others	Keeps the same ethical lesson accurately	No major difference
Overall judgment	Highly accurate as a whole text	Highly accurate as a whole text	Both translations achieve a similarly high level of whole-text accuracy

Suggested score

- Prompted translation: **5/5**
- Non-prompted translation: **5/5**

Table 5. (Story 2) Comparison of the Prompted and Non-Prompted Translations in Terms of Stylistic Suitability for Children

Criterion	Prompted Translation	Non-Prompted Translation	Difference
Vocabulary simplicity	Uses clearer and more accessible wording	Uses more formal and abstract vocabulary	Prompted version more child-friendly
Sentence readability	Sentences are smoother and easier for children to follow	Sentences are more formal and denser	Prompted version more readable
Narrative tone	More story-like and engaging	More rhetorical and expository	Prompted version more engaging
Accessibility of moral lesson	Moral idea is easier for children to understand	Moral idea is preserved but expressed more heavily	Prompted version better suited to children

Criterion	Prompted Translation	Non-Prompted Translation	Difference
Emotional clarity	The lioness's grief and realization are more immediate	Emotional meaning is present but less direct	Prompted version stronger
Suitability for ages 8 to 12	More appropriate for younger readers	Less suitable because of formal diction	Prompted version clearly better
Overall judgment	Highly suitable for children	Moderately suitable, but not ideal for children	Prompted version clearly superior

Suggested score

- Prompted translation: **5/5**
- Non-prompted translation: **3/5**

Table 6. (Story 2) Textual Evidence Supporting the Comparative Analysis

Source Meaning	Prompted Translation	Non-Prompted Translation	Analytical Observation
الظلم والعدوان	"harming others"	"committing injustice and aggression"	The prompted version uses simpler vocabulary
أثر العقاب	"when the punishment finally comes"	"before the evil consequence descends"	The prompted version communicates the idea more directly
العبرة الأخلاقية	"as you judge, so shall you be judged"	"as you judge, so shall you be judged"	Both preserve the moral lesson accurately
حزن اللبؤة	"She screamed, cried out, and rolled over and over in grief"	"She screamed, cried out, and rolled over on her back and belly"	The prompted version creates a more vivid emotional image
نصيحة الشعهر	"Bear patiently from others what others once bore patiently from you"	"Endure from others in the same way that others once endured from you"	The prompted version sounds more natural and child-friendly

These examples show that the prompted translation regularly provides moral and emotional content in a more explicit and accessible language, while retaining the original narrative structure and ethical message. The unprompted version is still semantically correct, but it is more reliant on formal, abstract, and rhetorically dense expressions. Hence, the prompted translation is more successful in conveying both the events and the moral lesson to the readers of 8–12 years. This supports the finding that prompting is more strongly associated with audience-orientation and readability than with basic semantic preservation.

4.5 Concluding comparison

The comparison of the two English versions of Bab al-Labwah wa al-Shihr shows that both translations are accurate at the whole-text level of meaning, because both translations preserve the source plot, the moral movement of the story and the final ethical lesson . But again, the prompted translation does a better job in two respects: it conveys the meaning more clearly in English, and it adapts the story more successfully for children. The translation without prompt is still correct, but it is more literal, more formal and less accessible for younger readers. Therefore, this second story confirms the same tendency as the first story: prompting improves stylistic suitability for children more strongly than it changes the global accuracy of the translation.

4.6 Analysis of the Prompted and Non-Prompted Translations of *Bab al-Nasik wa Ibn 'Irs*

In both translation conditions, the overall semantic and thematic structure of Bāb al-Nāsik wa-Ibn ʿIrs is preserved successfully. In both outputs, the core narrative landmarks are completely intact: the king requests a moral example, the philosopher sets the framework, the text describes the wife's warning against planning without certainty, the embedded anecdote of the ascetic daydreaming over his jar of butter and honey, the birth of the child, the faithful weasel killing the snake to protect the infant, the ascetic's rash misinterpretation of the blood, his hasty killing of the weasel, and his final, crushing regret about acting without verification. The difference between the two versions is only vocabulary and style, so that one is more or less accessible to young

readers than the other. Bāb al-Nāsik wa-Ibn ʿIrs . The human evaluation scores for the whole-text translation accuracy are shown in Table 7 . The comparative ratings for the stylistic suitability for children aged 8-12 are shown in Table 8 . The empirical scores suggest a high level of comparability in whole-text accuracy. While both models retain the two-part moral warning against speaking without certainty and acting without verification, However, the non-prompted translation still relies on formal and text-bound expressions, taken verbatim from the classical Arabic source. The expressions “performs an act without deliberation and without careful verification”, “provisions of barley meal”, “discipline him strictly”, “his mind was stripped away” are correct, but sound clinical, heavy and too abstract for English children’s literature. By contrast, the audience-challenged version translates these concepts into smoother, child-oriented prose, as described in the textual evidence in Table 9. The prompted variant lowers conceptual complexity by selecting familiar lexical items. Bi-ghayr ruwiyyah is intuitively translated as "without careful thought" as opposed to "without deliberation". Al-tathabbut is intuitively translated as "making sure first" as opposed to "careful verification"The prompted model also enhances narrative flow and emotional resonance significantly. The prompted version of the well-known jar story has a more natural storytelling rhythm (he kept calculating like this for five years) than the less natural, literal version of the non-prompted one (he continued calculating in that manner for five years). What is more, when describing the ascetic's abrupt shock at discovering his mistake, the prompted model changes the formal language "his mind was stripped away" to the emotionally immediate and accessible "he lost his senses". These findings confirm that the default behaviour of AI can preserve what takes place in a narrative, but an audience-oriented prompt is needed to transform a rhetorically dense classic in Arabic into a compelling, readable text for children.**Table 7. (Story 3) Comparison of the Prompted and Non-Prompted Translations in Terms of Translation Accuracy as a Whole**

Criterion	Prompted Translation	Non-Prompted Translation	Difference
Preservation of main events	Preserves all major events of the source story	Preserves all major events of the source story	No major difference
Preservation of inserted anecdote	Retains the story of the butter and honey jar fully and clearly	Retains the same inserted story accurately	No major difference
Preservation of moral structure	Maintains the warning against speaking without certainty and acting without verification	Maintains the same two-part moral structure	No major difference
Preservation of narrative sequence	Maintains the progression from the introductory warning to the final moral lesson clearly	Maintains the same narrative progression	No major difference
Faithfulness to final lesson	Keeps the lesson about haste, regret, and lack of verification clearly	Keeps the same lesson accurately	No major difference
Overall judgment	Highly accurate as a whole text	Highly accurate as a whole text	Both translations achieve a similarly high level of whole-text accuracy

Suggested score

- Prompted translation: **5/5**
- Non-prompted translation: **5/5**

Table 8. (Story 3) Comparison of the Prompted and Non-Prompted Translations in Terms of Stylistic Suitability for Children

Criterion	Prompted Translation	Non-Prompted Translation	Difference
Vocabulary simplicity	Uses simpler and more accessible wording	Uses more formal and abstract wording	Prompted version more child-friendly
Sentence readability	Sentences are smoother and easier to follow	Sentences are more formal and denser	Prompted version more readable
Narrative tone	More natural as a children's story	More formal and translated in tone	Prompted version more engaging
Accessibility of moral lesson	Moral lesson is easier for children to grasp	Moral lesson is present but expressed more heavily	Prompted version better suited to children
Emotional clarity	The regret, shock, and tragedy are more immediate	Emotional content is preserved but phrased more formally	Prompted version stronger
Suitability for ages 8 to 12	More appropriate for younger readers	Less suitable because of formal diction	Prompted version clearly better
Overall judgment	Highly suitable for children	Moderately suitable, but not ideal for children	Prompted version clearly superior

Suggested score

- Prompted translation: **5/5**
- Non-prompted translation: **3/5**

Table 9. (Story 3) Textual Evidence Supporting the Comparative Analysis

Source Meaning	Prompted Translation	Non-Prompted Translation	Analytical Observation
بغير روية	"without careful thought"	"without deliberation"	The prompted version uses more familiar vocabulary
التثبت	"making sure first"	"careful verification"	The prompted wording is easier for children
قصة الجرة	"he kept calculating like this for five years"	"he continued calculating in that manner for five years"	The prompted version reads more naturally
رد فعل الناسك	"he lost his senses"	"his mind was stripped away"	The prompted version conveys the meaning more clearly
العبرة النهائية	"acts in haste and does not make sure first"	"acts in haste and does not make sure"	The prompted version reinforces the moral lesson more explicitly

The above examples clearly show that the prompted translation uses simple, natural and reader friendly language while respecting the narrative order and moral importance of the source text. However, the non-prompted version often uses formal, abstract language that may be less accessible to younger readers. Both translations get the gist of the story across, but the prompted translation is more readable and better at storytelling. This is further evidence of the general trend seen throughout the study that prompting has more of an effect on stylistic suitability and audience adaptation than on the retention of global semantic content.

4.7 Concluding comparison

The quantitative indicators reveal a uniform trend over the entire corpus. Prompting had almost no difference in the semantic preservation of the whole text, both experimental conditions being evaluated perfectly (5/5) . That means that the native structure and semantic powers of the default AI language behavior are sufficient to track macro-level plots, structural logic, and ethical consequences without any external assistance. However,

there is a clear division in the stylistic suitability metric, as the non-prompted baseline performed poorly with a moderate score (3/5), due to its continued formal, adult-oriented diction, whereas the audience-prompted variants always scored perfect (5/5).

Table 10 summarizes the comprehensive empirical results across all three analyzed text selections from *Kalīlah wa-Dimnah*.

Story	Prompted Accuracy	Non-Prompted Accuracy	Prompted Style for Children	Non-Prompted Style for Children
Bab al-Qird wa al-Ghaylam	5	5	5	3
Bab al-Labwah wa al-Shihr	5	5	5	3
Bab al-Nasik wa Ibn ʻIrs	5	5	5	3

The findings suggest that prompting had little effect on overall accuracy of the whole text, since both translation conditions consistently preserved the main events, narrative structure, character relationships, and moral lessons of the source texts. However, prompting led to a substantial increase in stylistic appropriateness for children. Prompting had the strongest effect on audience-oriented adaptation across all three stories rather than on the preservation of basic plot meaning.

4.8 Overall Results of the Three Stories

The synthesis of the results of the three stories analyzed shows a stable, highly predictable comparative pattern. In both cases, both translation conditions could preserve the major narrative arches, embedded internal anecdotes, and central moral frames of the source texts. Thus, there was no significant difference in the global semantic fidelity between Bāb al-Qird wa-al-Ghaylam, Bāb al-Labwah wa-al-Shihr and Bāb al-Nāsik wa-Ibn ʿIrs. The main difference between the two experimental conditions is only in stylistic realization and audience-oriented framing, as summarized in Table 11. The non-prompted baseline outputs fall back systematically on a formal, text-bound texture mirroring the classical rhetorical complexity of Ibn al-Muqaffaʿ's Arabic prose. This dependence on literal syntax leads to accurate but structurally dense translations, heavy and conceptually aliening to a child readership. On the contrary, the translations, triggered by the audience, systematically avoid this stylistic barrier. With intentional lexical simplifications, natural storytelling cadences, and clearer emotional clarity, the guided model improves readability for children 8-12 years old, without altering the meaning or deleting the structure. For me, this pattern across separate stories is very strong evidence that the stylistic shifts we are seeing are a direct result of the audience constraint and not some accidental or text specific phenomenon.

Table 11. General Comparative Pattern Across the Three Stories

Variable	General Result in Prompted Translation	General Result in Non-Prompted Translation	Comparative Outcome
Accuracy of translation as a whole	Consistently high; preserves major events, narrative structure, character relations, and moral lessons	Consistently high; preserves major events, narrative structure, character relations, and moral lessons	No substantial difference
Stylistic suitability for children	Consistently high; simpler vocabulary, clearer flow, and a more child-oriented tone	Moderate; understandable, but more formal and less accessible to children	Prompted version clearly superior

4.9 General Interpretation of the Results

Results suggest that an audience-directed prompt does not change the semantic core of the AI-generated translation, but rather reorchestrates how that meaning is expressed to a specific target readership. If no constraints are specified the AI systems tend to default to a neutral generalized or over formal style which is correct but practically insufficient. The user provides a working framework by embedding a well-formed

translation brief into the prompt structure that instructs the system on how to tailor lexical choice, narrative tone, and sentence readability to the cognitive and linguistic expectations of young readers. Thus, prompt engineering becomes a crucial intermediate layer in AI workflows, connecting the raw semantic processing with functional communicative efficacy in literary contexts.

5. Discussion

The empirical analysis results of the three selected stories from Kalīlah wa-Dimnah show a stable and quite predictable trend. The large language model exhibited an intrinsic structural capacity to maintain global semantic content, perfectly tracking principal plot points, character roles, and embedded moral frameworks under both audience-prompted and non-prompted experimental conditions. But there was a big difference in terms of functional adequacy and reader access. Both conditions gave similar numbers for whole-text accuracy. The audience-prompted versions always outperformed the non-prompted baselines for stylistic suitability for readers 8-12 by a large margin. These findings strongly support the functionalist paradigm of translation studies, particularly the Skopos Theory as formulated by Vermeer (1989) who proposed that the success of a translation is evaluated based on its ability to achieve the communicative purpose (skopos) in a target context. The user introduces a constraint mechanism by explicitly stating an audience-specific translation brief in the prompt architecture, thereby moving the system's focus of operation from literal equivalence to functional adequacy. This is consistent with Sharkas (2025) who observed that incorporating translation brief elements such as target readership and readability constraints into LLM prompts systematically guides language generation toward specific stylistic parameters. Furthermore, the stylistic choices made in the prompted condition lend support to Oittinen's (2000) argument that translation for a young audience is, in essence, a matter of conscious adaptation of the text, and not simply code-switching. Instead of shallow lexical substitutions, the audience-prompted model changed sentence readability, simplified dense syntax and created a natural storytelling flow. This confirms the observations of Pourkamali et al. (2024) who argued that strategic prompt configuration does not distort the underlying semantic core, but fundamentally reshapes its textual realization and ease of discourse processing. Crucially, these findings provide a context on the baseline operational behavior of unprompted AI models. As Ataman (2025) pointed out, modern LLMs have robust multilingual processing capabilities that automatically ensure a high level of semantic accuracy at the macro level in narrative translation tasks without explicit guidance. This accounts for the perfect accuracy scores of the non-prompted baseline. However, unprompted systems by default employ a neutral, highly formal and text-bound register that reflects the rhetorical complexity of the source text, compromising reader-centric accessibility . Thus, prompt engineering does not serve as a tool for correcting semantic errors but as a crucial functional layer that mediates between broad semantic accuracy and reader-specific pragmatic reception.

5.1 Implications for AI-Assisted Literary Translation

The results suggest that prompt design should be treated as an integral part of AI-assisted literary translation. Prompting provides translators a practical way to tailor translations to particular readerships. For teachers, it offers a useful tool for developing translated materials appropriate for different age groups. For AI developers, the results underscore the need for audience-sensitive generation mechanisms, beyond simple semantic accuracy, that promote communicative effectiveness.

6. Conclusion and Recommendations

6.1 Conclusion

This research examined the effect of audience-oriented prompting on AI translation of classical literary stories for children using a corpus of three selections from Kalīlah wa-Dimnah. The evaluation considered two important operational variables: translation accuracy of the whole text and stylistic appropriateness for young readers, 8–12 years old. The findings indicate that default, unprompted AI models can hold together macro-level plot structures, character dynamics and central moral lessons quite well, but automatically reproduce the formal, archaic syntax and dense rhetoric of the classical source prose, making the text less accessible to younger minds. But the AI can overcome stylistic challenges if the prompt interface has a specific translation brief. In the audience-prompted condition, the conceptual complexity was systematically reduced by simplifying the vocabulary, increasing emotional immediacy and ameliorating the narrative flow, while preserving the entire narrative and ethical framework of the original text. Finally, the study concludes that the prompt design is an important functional variable in the performance of LLMs. First of all, you have to prompt for the audience. To turn rhetorically dense literary classics into readable, engaging and functionally adequate translations, without loss of core meaning.

6.2 Recommendations

The obtained results indicate the future of research and professional practice:

- Required Audience Briefs: Children's literature for AI-driven translation pipelines must specify constraints on prompt in terms of target age, limits in language and tone expectations to be functionally useful.
- Multidimensional Evaluation Metrics: The evaluation of AI-assisted translation should go beyond simple sentence-level error counts and semantic equivalence, and should also include macro-level discourse parameters such as readability, stylistic consistency, and audience reception.
- Extension of the Analytical Corpus: Future work should pursue this line of inquiry by testing a wider variety of genres of literature, poetic texts, and large datasets to evaluate the cross-genre stability of audience-oriented prompting.
- Cross-Platform Systems Testing: Repeat the study on different LLM architectures to see if the stylistic changes observed in this paper are consistent across competing AI systems.
- Integration of Multi-rater Human Evaluation: Future work should involve multiple independent raters, like experts in child psychology, pedagogy and translation studies, to obtain inter-rater reliability metrics and reduce the bias from subjective assessment.

References

Ataman, D., Birch, A., Habash, N., Federico, M., Koehn, P., & Cho, K. (2025). Machine translation in the era of large language models: A survey of historical and emerging problems. *Information*, *16*(9), Article 723. https://doi.org/10.3390/info16090723

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., . . . Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, *33*, 1877–1901. https://arxiv.org/abs/2005.14165

He, S. (2024). Prompting ChatGPT for translation: A comparative analysis of translation brief and persona prompts. In *Proceedings of the 25th Annual Conference of the European Association for Machine Translation (EAMT 2024)* (pp. 262–271). European Association for Machine Translation.

House, J. (2015). *Translation quality assessment: Past and present*. Routledge.

Koehn, P. (2020). *Neural machine translation*. Cambridge University Press.

Lathey, G. (2016). *Translating children's literature*. Routledge.

Oittinen, R. (2000). *Translating for children*. Garland Publishing.

Pourkamali, N., & Sharifi, S. E. (2024). *Machine translation with large language models: Prompt engineering for Persian, English, and Russian directions*. arXiv. https://arxiv.org/abs/2401.08429

Rivera-Trigueros, I. (2022). Machine translation systems and quality assessment: A systematic review. *Translation Spaces*, *11*(1), 1–28. https://doi.org/10.1075/ts.21014.riv

Sharkas, H. (2025). Exploring the role of translation brief elements in prompts to large language models. *Arab World English Journal for Translation & Literary Studies*, *9*(1), 1–18. https://doi.org/10.24093/awejtls/vol9no1.1

Vermeer, H. J. (1989). Skopos and commission in translational action. In A. Chesterman (Ed.), *Readings in translation theory* (pp. 173–187). Oy Finn Lectura.

Yamada, M. (2023). Optimizing machine translation through prompt engineering: An investigation into ChatGPT's customizability. In *Proceedings of Machine Translation Summit XIX, Vol. 2: Users Track* (pp. 195–204). Asia-Pacific Association for Machine Translation. https://aclanthology.org/2023.mtsummit-users.19

Zappatore, M., Longo, A., & Bochicchio, M. (2024). AI-based translation in healthcare: Opportunities and risks. *Applied Sciences*, *14*(2), Article 482. https://doi.org/10.3390/app14020482