

***On the Assessment of Chat GPT-4 Translation Accuracy
between Arabic and English with Special Reference to
Polysemous and Culture-bound Items***

حول تقييم دقة ترجمة ChatGPT-4 بين اللغة العربية
والانجليزية بالإشارة إلى الكلمات متعددة المعاني والعناصر الثقافية

Wissam BENYAHIA*

Received: 18/12/2024

Accepted: 25/05/2025

Published:20/06/2025

Abstract

The underscored advancement of Open AI's large language model (ChatGPT) has earned immense attention in the field of translation. By assessing ChatGPT-4 translation performance, this study discusses the translation accuracy between the distant language pair of Arabic and English and probes the strengths and challenges raised by the distinctiveness of translating polysemous words and culture-bound references based on a human user experience evaluation. The study involves samples from various text typologies to maximize the validity of the findings. Subsequently, in the course of evaluation and quality assessment, the results indicate that ChatGPT-4 demonstrates high accuracy when translating polysemy in various contexts and resorts to literal translation when rendering a few cultural references such as idiomatic expressions and proverbs. The study suggests further fine-tuned training to enhance translation consistency and performance, particularly in regards to unrelated

w.benyahia@univ-blida2.dz

language pairs such as Arabic and English to assure language transfer accuracy and offer insights for more extensive research on other complex issues of language.

Keywords: Translation Accuracy; ChatGPT-4; Polysemy; Cultural References; Translation Quality Assessment.

ملخص

حظي التقدم الملحوظ للنموذج اللغوي ChatGPT المستحدث من شركة OpenAI باهتمام كبير في مجال الترجمة، فمن خلال تقييم أداء ترجمة ChatGPT-4، تناقش هذه الورقة البحثية دقة الترجمة بين الثنائيات اللغوية الغير مترابطة مثل اللغة العربية و الإنجليزية، حيث تتقصى التحديات التي تثيرها عملية نقل الكلمات متعددة المعاني والعناصر الثقافية استنادا إلى تقييم تجربة المستخدم البشري. تتضمن هذه الدراسة عينات من مختلف أنماط النصوص لتعزيز صحة النتائج المتحصل عليها، وفي سياق تقييم جودة الترجمة، أشارت النتائج إلى أن ChatGPT-4 أظهر دقة عالية عند ترجمة الكلمات متعددة المعاني في سياقات مختلفة في حين انه لجأ إلى الترجمة الحرفية عند تقديم بعض العناصر الثقافية مثل التعبيرات الاصطلاحية والأمثال. اقترحت الدراسة تكثيف تدريب هذا النموذج لتعزيز اتساق الأداء الترجمي، خاصة فيما تعلق بالثنائيات اللغوية غير المترابطة لغويا وثقافيا مثل اللغة العربية والإنجليزية لضمان دقة عملية النقل اللغوي، كما تقدم الدراسة رؤى مستقبلية لمزيد من الدراسات المكثفة حول قضايا لغوية معقدة أخرى.

كلمات مفتاحية: الدقة الترجمية، العناصر الثقافية، ChatGPT-4 (شات جي بي تي-4)، تعدد المعاني، تقييم نوعية الترجمة.

Corresponding Author: Wissam BENYAHIA w.benyahia@univ-blida2.dz

1. Introduction

Translation, as a multifaceted medium of communication, has significantly evolved in the sphere of technological advancements. Such constant evolution has been notably reflected in the emergence of artificial intelligence language models that shaped how languages are approached and transferred. More fundamentally, chatbots such as Chat-GPT are developed and trained to understand, answer, and generate human-like discussions, questions, and texts. Among the various abilities natural language processing models can handle is the ability to perform translation tasks. Thus, the present paper presents a focused discussion on the accuracy of conveying polysemous words and culture-bound references between unrelated languages such as Arabic and English through a human user experience. Besides, as an evaluation of translation accuracy requires a systemic approach, the present paper invokes a descriptive-analytical method for an informed analysis of such a study whose primary objective is to unravel the quality of translation this version of natural language processing models (i.e. ChatGPT-4) can provide in light of the challenging language aspects when delivering translation tasks between Arabic and English.

Highlighting the particularities of ChatGPT, its strengths and challenges is intended to feed into the implications of such a model especially when performing complex tasks such as

translation. Extending the focus to encompass the translation of polysemy and cultural references is intended to fathom the translation quality and validity. Gearing the study towards the evaluation of translation between Arabic and English is intended to narrow down the focus to be on linguistically and culturally distant language pairs to test the translational decisions made. In all this, this paper invokes the latest and most recent literature as epitomized by researchers in the fields of focus.

Translation, as an interjective practice where various aspects of language are examined, is strongly founded on the analysis of textual intricacies and unravelling the implicit challenges of language. Accordingly, the process of transfer has rendered translation a complex and challenging task due to the multifaceted dimensions it covers when dealing with various languages. Thus, with the escalating demands for a comprehensive understanding of language and information as well as the urgent need to overcome barriers and challenges raised by linguistically and culturally unrelated languages, artificial intelligence was brought as a means of reconfiguring the industry of translation.

Interestingly, the increasing worldwide access to technology has paved the way for burgeoning interconnectedness across languages. Language translation has marked notable advancements in light of seamless communication with the integration of artificial intelligence. According to Mohamed et al, “the emergence of Artificial Intelligence (AI) in linguistic translation has revealed novel possibilities portending to span communication chasms and enable more fruitful cross-cultural connections” (2024, p.25553). One may then extrapolate that

such evolution has fostered translation as a medium of linguistic and cultural transfer and has become more indispensable than previously.

Despite the massive attention to various professional translation systems that were developed to automatically translate text materials from one language into another with the aid of software specifically designed for such a process, in recent years, there has been a rising interest in natural language processing models. More specifically, integrating large-scale pre-trained language models has gained an unprecedented trend in developing systems that incorporate translation tasks (Yang et al., 2019; Brown et al., 2020). Thoroughly elaborated, Gao et al (2023) hold that such language models are “trained on massive monolingual data, which allows them to capture rich representations of the input text” (2023, p.1). Eventually, such language models have shown invaluable capabilities in natural language processing tasks including ones that involve translation.

Among the famous and widely spread models that generate natural language responses based on a given input (i.e. prompts) is Chat GPT. Developed and introduced by Open AI, this chatbot uses generative artificial intelligence through a deep neural network that consists of many layers of transformers. To illustrate, in the process upon which Chat GPT is trained, Kalla, Smith Kuraku and Samaah (2023), in a paper that tackles the implementation and working of Chat GPT, state that “a large corpus of text data is fed into the model, allowing it to learn patterns and relationships between words, phrases and

sentences” (p.827). Following such a deep process of training, Chat GPT is then considered fined-tuned to conduct various tasks, including answering questions, generating content, and translating multiple textual materials.

Unsurprisingly, in the course of its evolution, ChatGPT has demonstrated remarkable advancements in performance throughout various versions starting from ChatGPT-1 up until the latest iteration ChatGPT-4 which was released on March 2023. According to the Open AI (2023) technical report that provides an in-depth look at the most salient features of this upgraded version, ChatGPT-4 is proven to: a) handle sophisticated queries; b) respond to complex tasks in language generation and comprehension; and c) demonstrate better abilities across a wide array of benchmarks. Such improvements paved the way for ChatGPT-4 to exhibit human-level performances at various scales. Nevertheless, Artetxe, Labaka, Agirre, and Cho (2018) hold that Chat GPT displays a ‘non-negligible’ performance gap when comparing it to other trained translation systems, such as Google Translate. Accordingly, in the interest of evaluating this large natural language processing model, this paper presents a focused examination of the translation accuracy of ChatGPT-4 from a human user experience, when translating language-specific items, such as polysemous words and culture-bound references between unrelated languages like Arabic and English.

2. Literature Review

2.1 Approaches to Translation Accuracy

There has been a wide array of perspectives on the particularities of practicing translation that serve as a strong basis upon which translation accuracy is achieved. Given that the multifaceted issues inherent in translation have always been challenging, the issue of equivalence remains problematic above them all. According to Biguenet and Schulte (1989), “an exact equivalence from one language to another will never be possible. This could be characterized as both the dilemma and the challenge for the translator” (p. xiii). To elaborate, in delving into the issues of equivalence, they seem to cover two major dimensions; cultural and linguistic patterns. Furthermore, the process of translation becomes complex if, supposedly, the linguistic aspect is settled and the cultural aspect is not undertaken. In this regard, the circulating translation theories and strategies that reflect how equivalence is approached function pursuant to two main extremes, notably, direct translation (i.e. fidelity) and free translation. More specifically, the former holds the translator invisible in the process of translation and the latter permits the translator to recreate meanings relying on various influencing factors, such as motives, agendas, competence and knowledgeability.

The existing literature that draws on the development of translation theories and approaches is extensively comprehensive in its coverage; it varies based on the extent to which translation can deal with the notion of equivalence along with the degree of variations between languages and their represented cultures.

Translation equivalence has been conceptualized in a paradigm of opposite, but interrelated, poles according to the stance of various approaches, namely, linguistic, sociolinguistic, functional and cultural approaches. From a linguistic stance, the translation aim is to convey meanings through a set of linguistic signs and symbols. Among the prominent advocates of this approach is John Catford (1965), whose contribution differentiates between ‘formal correspondence’ and ‘textual equivalence’. The sociolinguistic approach, however, endeavors to address the interplay between translation and communicative theories. This approach does not fully neglect the linguistic structures in the process of language but draws attention to the wide spectrum of external factors (henceforth, sociolinguistics) that influence translation. In the realm of this approach, Eugene A. Nida, in his book *Towards a Science of Translating* (1964), suggests two systemic orientations, namely, ‘formal equivalence’ and ‘dynamic equivalence’. It could be argued that Nida (1964) has brought new insights on breaking into translation equivalence by focusing on recreating the same impact of the source text on the target text readers.

In light of the incremental developments of translation theories, the functionalist approach was introduced with notable publications of theorists, such as Katharina Reiss (1976), Hans Vermeer (1978), Justa-Holz Manttari (1978) and Christian Nord (1997). This approach developed its perception of equivalence as a reflection against the paradigm of the linguistic approach. Subsequently, considering the various elements involved in the process of transferring texts from one language into another, translation approaches have witnessed new insights with the emergence of the cultural approach. Such a new turn in

approaching translation was marked by an emphasis driven by the influence of the cultural aspect of language. Various contributions that shed light on the issues of culture have received remarkable recognition in the field of translation. Such acknowledgement of the cultural approach is seen in Lawrence Venuti's *The Translator's Invisibility* (1995) which handles the cultural aspect of translation by distinguishing two main procedures, namely domestication and foreignization.

It has become clear from various overviews obtained on the nature and developments of translation approaches that translation accuracy serves as a meticulous element that reflects translation quality, particularly in addressing pressing challenges such as translation equivalence. Ultimately, by recognizing the perceptions gained on the significance of translation accuracy, it could be extrapolated that translation approaches are considered a testament to the complexity of translation acts and that the translation process is not a matter of peripheral concern to transferring words from one language into another. Rather, it is a holistic process that encompasses linguistic and extra-linguistic elements of language.

2.2 Navigating Culture through Translation

Translation is considered a multifaceted means of communication in which culture is set to be a cornerstone that reflects the complex nature of translation practice. Delving into the particularity of cultural differences has caught the attention of various theorists in the field of translation. Many contributions of theorists have brought to the fore such dimensions to discuss cultural discrepancies and their variations. Such contributions

include those of Eugene Nida (1964/2000), Mona Baker (1992), and Ali Almanaa (2016) among many other theorists. According to Eugene Nida (2000), “the ideas represented in the source text are a reflection of a particular culture, and the intended audience of a translation may not share such concepts nor replace the same value on the corresponding ideas” (p. 6). One may then argue that what is expressed and accepted in one culture might not be so in another.

More fundamentally, all languages are representations of different cultures they belong to and, eventually, such languages are loaded with cultural references in which they are labeled as ‘culture-specific’. The degree of complexity of translating such items varies according to many reasons. To elaborate, the context of the source text may contain cultural elements (i.e. words or expressions) that represent an overview of a different value or ideology than that of the target text. In this case, the text can be culturally distant to translate and may subsequently lead to translation loss, under-translation, mistranslation, or over-translation.

Accordingly, in the same line of thought, Eugene Nida (1964) holds that “the person who is engaged in translating from one language into another ought to be constantly aware of the contrast in the entire range of culture represented by the two languages” (p.20). This contrast is highlighted in the case of bi-directional translation between Arabic and English. For instance, there is a wide array of words and expressions in Arabic, such as الإحسان and الإيثار that reflect distinct meanings in which it is hard to find a direct equivalent that represents the exact meaning due to the absence of such notions in non-Muslim English-speaking

communities. Here, the translation would inevitably, opt for a gloss or paraphrase the item to minimize translation loss.

2.3 Polysemy: Shades of Meaning Multiplicity

Notably, while cultural references pose many challenges in translation, polysemy stands as a representation of the complexity of languages in the process of rendering meanings from one language into another. The notion of meaning ‘multiplicity’ is the most fundamental aspect of understanding and defining polysemy, the concept which scholars bring to the fore when defining this notion in translation. In his book *Linguistic Terms and Concepts*, Geoffrey Finch (2000) defines Polysemy as “a sense relation in which a lexeme has acquired more than one meaning” (p.173). This echoes what Sebastian Löbner (2002, p.44) draws that “polysemy plays a major role in the historical development of word meaning because lexemes continually shift their meanings and develop meaning variants”. In his elaboration, “polysemy is abundant and it is rather the rule than the exception and a lexeme constitutes a case of polysemy if it has two or more interrelated meaning variants” (Sebastian Löbner, 2002, p.44). Thus, polysemy can be inferred as a particularity of language in which words are more likely to have various meanings in different contexts of situations.

The co-occurrence of polysemous words can appear at different levels of language, and it can be seen in simple sentences, up to specialized terms. In Arabic, a huge number of words are considered polysemous, some of which have unrelated meanings. The word *أُتِي*, for instance, has multiple contextual

meanings as follows: *come* as in *أتى الرجل من مكان بعيد*, *grant* as in *يؤتي المرء حقه*, and *commit lewdness* as in *إتيان المعصية*. With all these considerations in mind on the nature of polysemy, its scope, and its place in the translation process, it could be extrapolated that the lack of contextualization is problematic that, given the various meanings a word can hold in different contexts, ambiguity is inevitable and, eventually, translation loss is highly to occur.

3. Methodology

As this study purports to assess ChatGPT-4 translation accuracy between Arabic and English in addressing polysemous words and culture-bound references, the methodology adopted focuses on evaluating the quality of translation at various levels. This includes addressing semantic, stylistic and functional accuracy in the target text. In light of the development of assessment systems in the realm of machine translation and large language models, human evaluation still provides more comprehensive and accurate feedback. Thus, to investigate the potential of ChatGPT-4 translation performance, this study employs human evaluation rubrics to test the accuracy of the translation. These rubrics assess whether the meaning of the polysemous words and cultural references is accurately conveyed in the translation, including the correctness of grammatical structures, vocabulary, and consistency of language.

3.1 ChatGPT-4 Prompt Design

Large language models, such as ChatGPT-4 are trained to generate responses based on commands that are commonly referred to as ‘prompts’. The design of prompts highly

contributes to triggering the translation ability of large language models. According to Yuan Gao, Ruili Wang and Feng Hou (2023), in their reflection on generating responses based on given prompts, “this nature determines that it is highly sensitive to both the content and form of the prompts” (p. 3). Yet, Jio et al (2023, p.3) hold that a similar translation performance could be attained by using different prompts. In their study, ChatGPT-4 provided three various prompts that would produce a highly approximate translation outcome. The prompts are as follows:

- TP1: Translate these sentences from [SRT] to [TGT]
- TP2: Answer with no quotes. What do these sentences mean in [TGT]?
- TP3: Please provide the [TGT] translation for these sentences

As ChatGPT-4 has been proven to be more flexible to variations in instructions and the performance is not overly reliant on the specific wording of prompts but the desired task, this study selects the following prompt in the course of analysis:

- Please provide an English translation for the following sentences in Arabic:
 - a. [Sentence X]

This prompt is a suggested variation of the aforementioned prompts given by ChatGPT-4. It is direct and clear in request and specifies the source/target languages that would guide ChatGPT-

4 to perform clear, less ambiguous translations and respond straightforwardly.

3.2 Sampling

The examination of ChatGPT-4 translation accuracy between Arabic and English is obtained by selecting a collection of passages and sentences from as varied texts as possible. The excerpts are therefore analyzed following the aforementioned evaluation rubrics of accuracy. Besides, the present study does not claim, by any means, to conduct an extensive analysis of all aspects of the language of Arabic and English. This would, subsequently, exceed by far the scope of this paper. To avoid drawing on personal intuition, the selection of excerpts intended for translation and examination is delimited to polysemous and cultural references. Thus, even though the delimitations of evaluating cultural references and polysemous items are only those posed by the subject under study, the analysis is deemed to be representative. These two references have been intentionally selected for evaluation as they raise critical challenges in the process of translation. Therefore, the study would confer credibility upon its findings.

4. Analysis

4.1 Polysemy

The following sentences were chosen from various contexts that range from political, sociological, literary and religious typology that contain the polysemous word ضرب:

- إن الشعوب إذا صارت لها الغلبة والقوة، فإنها تضرب على الناس بسيف السلطان،

وتقيم الحكم بما يقتضيه الدين والسياسة.

On the Assessment of ChatGPT-4 Translation Accuracy between Arabic and English with Special Reference to Polysemous and Culture-bound Items

- قال الله تعالى: ﴿فَضَرَبْنَا عَلَىٰ آذَانِهِمْ فِي الْكَهْفِ سِنِينَ عَدَدًا﴾. القرآن الكريم، سورة الكهف، الآية 11

- كان المعلم كلما رأى الصبي يُخطئ في قراءة النص، يضرب مثلاً من محفوظاته القديمة ليعلمه كيف يتقن اللغة.

The aforementioned sentences were introduced altogether to ChatGPT-4 following the selected prompt for study. It is worth mentioning that inserting the sentences all at once is intended to test how ChatGPT-4 can differentiate between the various meanings of the word ضرب in different contexts at the same time. Thus the sentences were introduced as follows:

- Please provide an English translation for the following sentences in Arabic:

- إن الشعوب إذا صارت لها الغلبة والقوة، فإنها تضرب على الناس بسيف السلطان، وتقيم الحكم بما يقتضيه الدين والسياسة.

- قال الله تعالى: ﴿فَضَرَبْنَا عَلَىٰ آذَانِهِمْ فِي الْكَهْفِ سِنِينَ عَدَدًا﴾. القرآن الكريم، سورة الكهف، الآية 11

- كان المعلم كلما رأى الصبي يُخطئ في قراءة النص، يضرب مثلاً من محفوظاته القديمة ليعلمه كيف يتقن اللغة.

The translation was rendered as follows:

- When a people gain dominance and power, they **impose their authority** upon others with the sword of sovereignty, and

establish governance according to what religion and politics require.

- Allah Almighty said: “So **We covered** their ears in the cave for a number of years.” (The Qur’an, Surah Al-Kahf, verse 11).

- Whenever the teacher saw the boy make a mistake in reading the text, he would **give an example** from his old memorized lessons to teach him how to master the language.

4.2 Cultural References

The following sentences contain elements or expressions that reflect the diverse nature of cultural references that are deeply embedded in the Arabic language and culture. The passages were introduced to ChaGPT-4 following the selected prompt for study:

- Please provide an English translation for the following sentences in Arabic:

- لقد بلغ السيل الزبي. لم يعد هناك مجال للصمت أو الانتظار، فكل تأخير سيعني خسارة أكبر وعليّ أن أواجه العواقب بكل شجاعة، مهما كانت.

- عندما تلقيت الخبر السار عن نجاح زملائي في العمل بعد جهود طويلة وشاقة، أثلج صدري أن أرى عملهم يُثمر ويُقدّر، فقد كانوا يستحقون ذلك

The translations were rendered as follows:

- **The flood has reached its limit.** There is no longer any room for silence or waiting, as every delay will result in a greater loss. I must face the consequences with courage, no matter what they may be.

- When I received the good news about the success of my colleagues at work after long and hard efforts, **it warmed my heart** to see their work bear fruit and be appreciated, as they truly deserved it.

5. Results and Discussion

5.1 Polysemy

The first sentence stresses the significant role of unity and establishing governance and authority. Here, the word ضرب means to enforce control upon people. In the suggested translation of ChatGPT-4, it could be notably seen that the word ضرب has been rendered as ‘impose authority’. This rendition has successfully conveyed the intended idea of exercising dominance and authority by force.

In the second sentence, the Quranic verse of Surat Al-Kahf describes the act that Allah has caused the companions of the cave to enter into deep sleep so that they stay isolated and safe from external harm. The word ضرب here means to seal or prevent the sense of hearing noise or getting disturbed. In the translation, ChatGPT-4 has suggested rendering this meaning by using the term ‘to cover’ to explicitly convey the intended meaning of the word ضرب in this context of situation.

The third sentence reflects the use of methods in the teaching profession. The narrative depicts such a teaching method in which the teacher uses examples that would help the boy grasp and understand language. The word ضرب in this

sentence has been used in describing the act of providing examples that would enhance language acquisition. The translation in English has succeeded in suggesting the equivalent needed as the context in which the word ضرب occurred has been supported by the element (مثالا) that has helped ChatGPT-4 select the exact equivalent. Without such support of surrounding words in the sentence, ChatGPT-4 could have confused with other meanings of the word ضرب, such as ‘hit’, ‘strike’, ‘turn’, ‘roam’, or ‘travel’ among other meanings.

One may clearly extrapolate that ChatGPT-4 has succeeded in producing a target language translation that bears a close meaning to the source language input sentences by taking into consideration the polysemous words when translating from Arabic into English. ChatGPT-4 could not only identify variant meanings of the polysemous word ضرب but gave great attention to the context and text typology as well, as they are highly significant in distinguishing the word meaning. This reflects ChatGPT-4 capability to provide human-like translations respecting the structure, grammar and fluency of language, taking into account the challenging aspects of language as in the case of polysemy.

5.2 Cultural References

The suggested sentences contain idiomatic expressions that are specific and widely common to the Arab culture. The first sentence demonstrates the symbolism of water and floods in the Arab culture in times of inevitable and uncontrolled accumulations; it expresses how the situation has reached a

crucial point where a critical decision is required. The expression *بلغ السيل الزبي* is also used in Arab communities to denote extreme frustration or urgency; it refers to the time that has passed for patience. In the context of this sentence, the expression similarly reflects the same meaning where the situation has escalated to a point where there is no longer space or room for tolerance or delay.

In the translation, it appears that ChatGPT-4 has opted for a literal translation for rendering the expression into English where the suggested translation does not reflect the stylistic and idiomatic form as in the source language. There exist similar equivalent expressions in English that highly denote such a meaning, like *enough is enough*, *a point of no return*, or *the last straw* that ChatGPT-4 could have used to convey the meaning in an idiomatic way. Yet, it could be argued that the translation did not deviate or alter the original meaning of the source language and rendered it via literal translation.

As for the second sentence, the context describes a situation of receiving good news for the success of colleagues after hard work. Such news conveys a sense of happiness and joy symbolized by the expression *أثلج صدري*. The Arabic language of Arab communities makes use of words and expressions that are inspired or influenced by the environment of their settlement. The expression *أثلج صدري* is used by Arabs (especially those who live in hot atmosphere) to depict the state of cooling the chest after drinking cold water in hot weather. The translation provided by ChatGPT-4 notably uses an equivalent idiomatic

expression used in the English language and culture to reflect the same meaning.

The expression *warmed my heart* implies the feeling of immense joy in response to good news. This expression is deeply rooted in English-speaking cultures, and it uses imagery of real-world climate to describe the cold weather of such cultures. Thus, it could be argued that the translation has, in one case succeeded in suggesting an expression that is different in form but reflects the same meaning in the same idiomatic sense and, in the other case, resorted to a literal translation with no extra affords to look for idiomatic expressions that deliver approximate meanings in English. This stance of translating cultural references differently could be justified by the insufficient training of ChatGPT-4 on a large database of corpora. Thus, ChatGPT-4 could not predict or suggest a similar expression respecting the idiomatic form as in the source text.

6. Conclusion

The present study delves into the realm of ChatGPT-4 translation accuracy between Arabic and English aiming to analyze polysemous and culture-specific items embedded in Arabic. The qualitative analysis demonstrates that ChatGPT-4 translation performance highly correlates with that of human translators in certain dimensions. The translation provided by ChatGPT-4 conformed to the standards of accuracy at the level of language structure, word choice and fluency.

Among the salient conclusions drawn from the assessment reveal that the translations present a consistent ability with human translators in addressing polysemy in different contexts and text typologies. Translating polysemous words requires the

distinction of meaning and selecting equivalent terms pertaining to those meanings. It has been observed that ChatGPT-4 has succeeded in extracting the meaning of the same word in various contexts and, thus, suggests a suitable translation.

Conversely, translations obtained when rendering culture-bound references from Arabic into English were faithful to the source text. Yet, the results were still questionable. More explicitly, the analysis has shown that ChatGPT-4 translation is clearly oriented to the source text language structure or covertly translated, neglecting the wide range of suggestions that convey the exact meaning. This is due to the fact that most of the cultural references in Arabic are inter-textually linked to the Arabic heritage and culture. Thus, suggesting an equivalent would require an extra database of texts and fine-tuned training.

Thus, the results highlight that ChatGPT-4 still needs further training and optimization in producing accurate translations when dealing with cultural references like idioms and proverbs that are deeply rooted in Arabic. This study is significant in a way that it aims to assess the translation quality of ChatGPT-4 from a human user experience perspective and highlights translation discrepancies. This would help assist in enhancing ChatGPT-4 performance in translating between Arabic and English.

It is worth mentioning that further studies on evaluating ChatGPT-4 translation accuracy should not only be reduced or limited to the cultural aspects of Arabic and English languages but also assess the semantic and stylistic framework of those languages. Also, it has become clear that assessing ChatGPT-4

translation accuracy is based on understanding the context of the source text. Therefore, assessing translations in future studies should consider challenges of equivalence that affect intelligibility and intentionality. Besides, it is highly recommended that future studies focus on other aspects embedded in the Arabic-English language pair as they are linguistically and culturally unrelated, thus, highlighting more stressing issues in translation.

7. Bibliography

Artetxe, M., Labaka, G., Agirre, E., & Cho, K. (2018). Unsupervised neural machine translation. In *International Conference on Learning Representations (ICLR2018)*.

<https://openreview.net/forum?id=HkWj7z-A->

Biguenet, J. and R., Schulte. (1989). *The Craft of Translation*. Chicago: The University of Chicago Press.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., & others. (2020). *Language models are few-shot learners*. In *Advances in Neural Information Processing Systems (NeurIPS 2020)*. <https://arxiv.org/abs/2005.14165>

Catford, J. C. (1965). *A Linguistic Theory of Translation*. London: Oxford University Press.

Finch, G. (2000). *Linguistic Terms and Concepts*. New York: Palgrave Macmillan.

Gao, Y., Wang, R., & Hou, F. (2023). Unleashing the power of ChatGPT for translation: An empirical study. *arXiv preprint arXiv:2304.02182*.

Jiao, W., Wang, W., Huang, J.-t., Wang, X., & Tu, Z. (2023). *Is ChatGPT a good translator? A preliminary study*. arXiv preprint arXiv:2301.08745.

<https://arxiv.org/abs/2301.08745>

Kalla, D., Smith, N., Kuraku, S., & Samaah, F. (2023). Study and analysis of Chat GPT and its impact on different fields of study. *International Journal of Innovative Science and Research Technology*, 8(3), 827.

<https://www.ijisrt.com/IJISRT23MAR956>

Löbner, S. (2002). *Understanding semantics* (2nd ed.). Routledge.

Nida, E. A. (1964). *Towards a Science of Translating: With Special Reference to Principles and Procedures Involved in Bible Translating*. Leiden: E. J. Brill.

Nida, E. A. (2000). 'A Fresh Look at Translating'. In A., Beeby, D., Insinger and Presas, M. (Eds). *Investigating Translation* (pp. 3-12). Amsterdam and Philadelphia: John Benjamins.

OpenAI. (2023). *GPT-4 technical report*. arXiv.

<https://arxiv.org/abs/2303.08774>

Venuti, L. (1995). *The Translator's Invisibility*. London/New York: Routledge.

Y. A. Mohamed, A. Khanan, M. Bashir, A. Hakim H. M. Mohamed, M. A. E. Adiel, and M. A. Elsadig, (2024) “The impact of artificial intelligence on language translation: A review,” *IEEE Access*, vol. 12, pp. 25553–25579, doi: 10.1109/ACCESS.2024.3366802.

Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R. R., & Le, Q. V. (2019). *XLNet: Generalized autoregressive pretraining for language understanding*. In *Advances in Neural Information Processing Systems (NeurIPS 2019)*.

<https://arxiv.org/abs/1906.08237>