



## RESEARCH ARTICLE

Section: *Literature, Linguistics & Criticism*

## Translating food idioms across the English-Arabic language divide: A comparative study of human translators, ChatGPT, and Gemini

Sadam Megdadi<sup>1</sup>, Zakaryia Almahasees<sup>2</sup>, Ahmad Al-Harashseh<sup>2</sup>, Rafat Al Rousan<sup>2</sup>, Mohammad Anwar Al-Taher<sup>3</sup> & Islam Husienat<sup>4</sup>

<sup>1</sup>Department of Arabic Language and Literature, Faculty of Arts, Yarmouk University, Jordan

<sup>2</sup>Department of Translation, Faculty of Arts, Yarmouk University, Jordan

<sup>3</sup>Department of English Language and Translation, Faculty of Arts and Humanities, Applied Science Private University, Amman, Jordan

<sup>4</sup>Department of Paediatrics ICU, King Abdullah University Hospital, University of Science and Technology, Jordan

\* Correspondence: [moteansperth@outlook.com](mailto:moteansperth@outlook.com)

### ABSTRACT

Idiomatic expressions, especially those rooted in the universal domain of food, represent one of the most culture-embedded and translation-difficult forms of language. The current study aims to examine the translation of twenty English food idioms into Arabic and twenty Arabic food idioms into English by human professional translators, as well as two of the most prominent large language models (LLMs): ChatGPT (GPT-4o) and Google Gemini Advanced. A mixed-methods approach is used to evaluate the performance of all sixty translations of each idiom on four criteria: semantic equivalence (SE)<sup>1</sup>, pragmatic appropriateness (PA), cultural adaptation (CA), and natural fluency (NF), with a five-point scale for each. A composite translation quality score is derived for each translator/agent. The results show that human professional translators outperform the two LLMs on all criteria with a mean translation quality score of 5.00, compared to 3.83 for ChatGPT and 3.60 for Gemini. The results also show the greatest performance gap between human translators and the two LLMs is on cultural adaptation. The results of the current study have important implications for the pedagogy of translation, translation practice, and the development of culturally sensitive AI translation tools.

**KEYWORDS:** food idioms, English–Arabic translation, Arabic–English translation, idiom translation, ChatGPT, Gemini, human translation, cultural equivalence, pragmatic appropriateness

### Research Journal in Advanced Humanities

Volume 7, Issue 2, 2026

ISSN: 2708-5945 (Print)

ISSN: 2708-5953 (Online)

### ARTICLE HISTORY

Submitted: 20 December 2025

Accepted: 20 March 2026

Published: 08 April 2026

### HOW TO CITE

Megdadi, S., Almahasees, Z., Al-Harashseh, A., Alrousan, R. M., Al-Taher, M. A., & Husienat, I. (2026). Translating food idioms across the english arabic language divide:A comparative study of human translators, ChatGPT, and Gemini. *Research Journal in Advanced Humanities*, 7(2). <https://doi.org/10.58256/pf7v7z49>



Published in Nairobi, Kenya by Royallite Global, an imprint of Royallite Publishers Limited

© 2026 The Author(s). This is an open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## 1. Introduction

Language and food are arguably among the most basic organizing principles in human societies (Frye & Bruner, 2012). These are universal in their importance and infinitely varied in their cultural specification. It is thus not unexpected that food should provide one of the richest fields of idiomatic language in the world's languages. From English 'piece of cake' to Arabic 'حلومو زبخ' (bread and salt) (Alhasnawi, 2024), food idioms convey values, relationships, and morality in condensed and conventionalized ways that do not easily lend themselves to translation. In fact, to translate these idioms is not merely to translate words but to move between worlds (Garzone, 2017).

The focus of this study is the translation of food idioms between English and Arabic. These two languages are linguistically and culturally distant. They are not only from different language families, with English belonging to the Germanic family and Arabic to the Semitic family, but also have radically different morphologies and cultural, religious, and historical underpinnings. English food idioms such as 'bring home the bacon' or 'cook the books' reflect Anglo-European cultural practices and metaphorical traditions. Arabic food idioms such as 'حلومو زبخ' or 'قوالح طلقعب لكأ' reflect Islamic ethics, Bedouin hospitality, and classical literary conventions. The cultural differences between English and Arabic make them a particularly fertile ground for investigating translation challenges (Al-Rushaidi & Ali, 2017; Mohammed, 2025).

The research is inspired by a recent technological advancement with significant practical implications: the rapid deployment of large language model tools, such as OpenAI's ChatGPT and Google's Gemini, as translation tools used by individual users, students, professional translators, and institutions (Ferrag & Bentounsi, 2024). The tools' ability to express fluency in translation is significant, and their usability has made them popular alternatives or supplements to human translation. However, little is known about the tools' potential to address the pragmatically complex demands of idiomatic translation, especially between culturally distant languages (Fitria, 2025).

This study fills the existing research gap by conducting an empirical investigation into the performance of human translators versus AI tools on a dataset of forty food idioms, with twenty translated into Arabic from English, and another twenty translated into English from Arabic. The study aims to answer the following three main research questions:

1. What is the performance of human translators, ChatGPT, and Gemini like on the translation of English food idioms into Arabic?
2. How do they compare in the translation of Arabic food idioms into English?
3. What qualitative patterns of success and failure characterise AI translation of food idioms in this language pair?

The paper is divided into the following sections: Section 2 includes a literature review on the theory of idiom translation, machine translation, and AI translation, as well as the difficulties associated with the translation of the English language into the Arabic language. Section 3 includes the methodology, while Section 4 includes the findings. Finally, Section 5 includes the discussion, while Section 6 includes the conclusion.

## 2. Literature Review

### 2.1 Idioms, Food, and Culture

Idiomatic expressions have been referred to as 'frozen' or semi-fixed multi-word expressions whose meaning is not transparently derivable from the sum of their component parts (Keizer, 2016; Meso, 2023). Lontas (2019) and Boers (2014) suggest a continuum ranging from 'pure idioms' whose meaning is fully non-compositional (e.g., 'kick the bucket'), through 'semi-idioms' and 'figurative idioms' to more conventionalised expressions whose idiomaticity is relatively transparent. Food idioms, as a particularly rich subset of this larger category, have been a focus of special interest because of their ubiquity—every culture uses food metaphors—and their specificity: the particular foods and practices which produce their idiomaticity are particularly embedded in a given culture.

According to (Majić Mazul, 2017), In his study of food idioms through the theories of cognitive linguistics, it is suggested that food idioms are usually conceptual metaphors that map food and eating concepts to other conceptual domains such as knowledge, emotion, greed, deceit, and social relationships. While it is true that cross-linguistically and cross-culturally, food idioms share similarities in their metaphors, such as the

universal concept of shared bread symbolising trust and solidarity, as seen in Arabic ‘حلمو زبـخ’, and English ‘bread and salt’, it is not possible to translate food idioms word-for-word.

In the context of the Arabic culture, there are further layers of religious and moral connotation to the food idioms. For example, the idea of ‘rizq’ or provision, which has such a central role in the theological constructs of Islam as a divine gift, imbues the phrases such as ‘ءامسلا نم قزر’ or ‘شـريـعـلا تمـقـل’ with connotations which are completely absent from their most literal translation into the related English phrases. Alomoush (2015) provide evidence of the prevalence and richness of food imagery in Arabic proverbs and idioms, where the metaphor of the bread (‘زبـخ’) has a social, moral, and spiritual significance which goes far beyond the literal.

## 2.2 Translation Strategies for Idioms

The fundamental taxonomy of idiom translation strategies proposed by Baker (2018) lists four major translation approaches: “the use of a target language idiom of similar meaning and similar form,” “the use of a target language idiom of similar meaning but dissimilar form,” “translation by paraphrase,” and “the omission of the idiom if no equivalent exists.” Such translation approaches are seen to represent a balance, according to the formal versus dynamic equivalence proposed by (Nida, 2001), between formal and dynamic equivalences. In the case of food idioms, the culturally arbitrary nature of the food item (the culturally specific meaning of the word “bacon” for English speakers, for example) will often require a departure from formal correspondence under the dynamic equivalent translation principle.

The framework of comparative stylistics proposed by (Pym, 2014; Vinay & Darbelnet, 1995) offers a set of complementary means, especially that of ‘equivalence,’ which involves selecting a target language equivalent that has an equivalent situational effect to that of the source language. The opposition between ‘direct’ and ‘oblique’ procedures is also seen to apply productively to food idiomatic translation, with ‘literal’ or ‘borrowing’ procedures rarely yielding acceptable target language output, while ‘adaptation’ and ‘modulation’ procedures are more likely to yield equivalence.

Pym (2020) suggests a theoretical framework for translation decisions as a matter of risk management: translators select between strategies that maximize communicative success and minimize the risk of pragmatic failure. For food idioms, the risk of pragmatic failure is high when the cultural specificity of the source idiom lacks a functional equivalent in the target culture and the translator must decide to domesticate or foreignize the idiom (Gregov, 2018). Kemppanen et al. (2012) and Yang (2010) arguments on the importance of foreignization as a strategy of cultural resistance to domesticization are also noteworthy as a counterposition to the domesticization trend in economic translation practice.

## 2.3 Machine Translation and Large Language Models

The history of machine translation has been a gradual struggle with the problem of non-compositionality of idiomatic language. While rule-based machine translation systems (RBMT) have failed to handle idioms systematically because of the compositional logic of rules, statistical machine translation systems (SMT) have been able to improve performance by learning phrase-level alignments from bilingual corpora. However, the problem of cultural specificity and frequency of idioms has still been an impediment to the performance of statistical machine translation systems. The current state-of-the-art neural machine translation systems, using transformer models with attention mechanisms, have been able to attain near-human performance on fluent translation for high-resource language pairs such as English-French or English-German (Almahasees, 2021; Almahasees et al., 2021; Monti, 2015).

The advent of massive language models, trained on large heterogeneous corpora and fine-tuned using reinforcement learning from human feedback (RLHF), represents another qualitative leap. ChatGPT, developed on the GPT-4 architecture (OpenAI, 2023), and Google Gemini Advanced (Google, 2024) have the ability to engage in multi-turn conversations, perform reasoning, and produce texts in dozens of languages with style awareness. Unlike previous MT systems, they can be supplied with context and control information that influence the translation output, thus providing a level of pragmatic awareness that is not possible in traditional pipeline MT systems.

However, empirical studies have shown that there are still areas of weakness. (Kleidermacher & Zou,

2025) reported that GPT-4 was comparable to commercial MT systems on well-resourced languages but weaker on morphologically complex and culturally distant language pairs. (Behzad et al., 2024) reported systematic over-literalization in ChatGPT's processing of Chinese idioms (chengyu), which are translationally transparent but culturally opaque. (Ochieng et al., 2025) demonstrated that the design of the prompt has a significant impact on the quality of LLM translation, with prompts providing context resulting in significantly improved idiomatic translation. In the case of Arabic, (Alkhasawneh, 2026) reported high levels of pragmatic failure in LLM translation of Gulf Arabic colloquialisms, (Al-Otaibi et al., 2025) reported morphological errors and register discrepancies in both ChatGPT and Gemini Arabic translations.

#### 2.4 English–Arabic Translation: Specific Challenges

The English-Arabic language pair is one of the most challenging in terms of translation. At the linguistic level, Arabic's root-pattern morphology, agreement system, absence of an equivalent to the copula, and different marking of grammatical gender and number create systematic mismatches with English grammatical structures. The diglossic nature of Arabic, with its distinction between Modern Standard Arabic (MSA, fusha) and a range of colloquial dialects (ammiyya), adds a dimension of complexity not found in English: the choice of MSA or a particular dialect is a register-sensitive choice with social and communicative consequences (Almahasees, Albudairi, et al., 2025).

At the cultural level, the differences are just as significant. English idioms often make reference to sports, maritime history, industrial capitalism, and secular law, while Arabic idioms tend to make more reference to Islamic religious discourse, classical poetry, especially the mu'allaqat form, desert and agricultural life, and tribal social structure. The result is that, at times, it is not possible to identify an equivalent idiom in the target language that carries the same cultural and pragmatic weight, as in the case of paraphrase, adaptation, or creative equivalence (Almahasees, Al-Harashseh, et al., 2025).

Studies on Arabic MT have shown the cumulative impact of these issues (Mansour et al., 2014) showed that there is a substantial degradation in the performance of SMT systems when they are required to translate dialectal Arabic. (Al-Mannai et al., 2014) pointed out that even the best NMT systems have a high error rate in processing Arabic idioms and metaphoric expressions compared to literal ones. These issues are further exacerbated in the AI age due to the presence of training data biases, where MSA is greatly over-represented compared to the colloquial dialects in large text corpora, and food idioms, which are very common in colloquial Arabic, are under-represented in formal written texts.

#### 2.5 Translation Quality Assessment

Translation Quality Assessment (TQA) has been the subject of considerable methodological debate (House, 2014). While automatic approaches, e.g., BLEU (Papineni, 2002), METEOR (Banerjee & Lavie, 2005), and BERTScore (Zhang et al., 2019), are efficient, they are not sensitive to pragmatic, cultural, and register considerations. Human evaluation approaches, e.g., the Multidimensional Quality Metrics (MQM) framework (Lommel et al., 2014), and LISA evaluation models, though providing higher analytical resolution, require considerable human resources. For the evaluation of translated idioms, several researchers have proposed that evaluation criteria should extend beyond semantic accuracy to include pragmatic, cultural, and linguistic nativeness (Ahmed, 2024). In this study, a multi-criterion human evaluation method, in accordance with the discussions, has been used, which is discussed in Section 3.

### 3. Methodology

#### 3.1 Research Design

The research uses a mixed-methods comparative research design, integrating quantitative scoring on four criteria of translation quality with qualitative text analysis of illustrative examples. The research design takes an empirical, corpus-based paradigm (Laviosa, 2004; Toury, 1999), and a purposefully constructed corpus is used to produce data amenable to systematic comparison between the two types of translators. The integration of quantitative analysis and qualitative analysis allows for statistical comparison of overall performance as well as detailed explanation of the underlying mechanisms.

### 3.2 Corpus Design and Idiom Selection

The corpus comprises forty food idioms, consisting of twenty English idioms translated into Arabic and another twenty Arabic idioms translated into English. Food idioms were chosen as the focus domain for this study for several reasons: they are semantically cohesive and culturally significant as a subdomain of idiomatic language, they are common in both informal and formal writing, they vary in level of opacity ranging from relatively transparent to completely opaque, and they are known to cause difficulties for MT systems as identified in previous research (Ferrag & Bentounsi, 2024; Laviosa, 2004).

English idioms included those representing various food items (cake, beans, bacon, fish, eggs, nuts, salt, bread, cheese, apple), semantic domains (ease/difficulty, deception, anger, industry, commerce, communication, social relationship), and degrees of opacity. Arabic idioms included those representing classical and modern usage, food items such as 'زبـخ' (bread), 'حـلم' (Eser et al.), 'مـقـول' (morsel/bite), 'قـوالـح' (sweetness), 'مـاعـط' (Alhabahba et al.), reflecting the importance of bread and staple food items in Arabic idiomatic expressions. Idioms whose translation was fully standardized in major bilingual dictionaries, as well as idioms whose primary problem was grammatical rather than idiomatic, were excluded. The corpus was finally checked by two independent bilingual translation experts for representativeness, cultural authenticity, and appropriate level of difficulty, with some modifications made according to their suggestions.

### 3.3 Data Collection

Human translations were also generated by three professional translators who have at least ten years of experience in English-Arabic translation and postgraduate qualifications in translation or linguistics. Translators worked independently from printed prompts that provide each idiom in a short sentential context. They were asked to provide natural-sounding translations that appeal to the general, educated reader. Translators did not have access to the AI translations and made their evaluations without knowledge of the other translators' work. The AI translation responses were obtained by querying ChatGPT (GPT-4o, accessed through Open AI API, March 2025) and Gemini Advanced (Google, accessed March 2025). A standardized prompt was used for each of the AI systems, as follows: *Please translate the following idioms from English into Arabic and vice versa. Provide a natural, idiomatic translation that captures the meaning and tone of the original sentence.* The prompts are identical in form for all the idioms and translator systems, with no additional context or metalinguistic information provided, only the sentence context. A new, separate prompt was provided for each idiom, with no repetition of prompts to avoid possible contamination between responses.

### 3.4 Evaluation Framework and Scoring

The translated texts were assessed by two bilingual assessors (Rater A: L1 Arabic, C2 English; Rater B: L1 English, C1 Arabic) with postgraduate qualifications in translation studies. The assessors were unaware of the type of translation (human or AI) and assessed the translated texts in random order. Each translated text was rated from 1 (completely inadequate) to 5 (fully adequate) on four criteria:

1. Semantic Equivalence (SE) - the translation conveys the idiomatic meaning of the source expression.
2. Pragmatic Appropriateness (PA) - the translation is appropriate in the target language culture and situation.
3. Cultural Adaptation (CA) - the translation successfully adapts the cultural references of the source expression to the target culture.
4. Natural Fluency (NF) - the translated expression sounds natural and idiomatic in the target language.

The inter-rater reliability was calculated using Cohen's Kappa coefficient, which resulted in  $\kappa = 0.81$  (SE), 0.78 (PA), 0.76 (CA), and 0.83 (NF), all of which reflected substantial to almost perfect agreement (Landis & Koch, 1977). In cases where the raters did not agree by more than one point, the decision was made by a third bilingual rater (Arabic-English, C1-C2 level). A composite Translation Quality Score (TQS) was calculated by taking the arithmetic mean of the four criterion scores. Statistical analyses were carried out using one-way ANOVA with Tukey's HSD post-hoc tests ( $\alpha = .05$ ).

#### 4. Findings and Analysis

##### 4.1 Full Corpus: English Idioms Translated into Arabic

Table 1 below shows the list of twenty English food idioms, their human, ChatGPT, and Gemini translations into Arabic, and the TQS scores.

*Table 1: English Food Idioms — All Translations into Arabic with TQS Scores (1–5)*

#	English Idiom	Human Translation (Arabic)	ChatGPT Translation	Gemini Translation	H	C	G
1	Piece of cake	أدج لهس	أدج لهس	ةياع يف رمأ قلوهسلا	5	5	4
2	Spill the beans	رسلا يشفي	رسلا فشكي	نع حصفي تامول عملا ةيرسلا	5	4	3
3	Butter someone up	اصخش قلمتي	اصخش قلمتي	اصخش حدمي هاضر بسكل	5	5	4
4	Cry over spilled milk	مدنلا نم ءدئاف ال ناوالا تاوف دعب	تاوف دعب مدنلا ناوالا	ءيش يلع نزلحلا حالص! نكمي ال	5	4	3
5	A hard nut to crack	قبعص قلكشم لحلا	لحلا بعص رمأ	تدق عم ءيضق	5	4	4
6	Bring home the bacon	تمقل بسكي شي عل	لاملا رفوي قرسألل	لاملا بسكي ةلئاعلا ءل اعلا	5	3	3
7	Go bananas	هباصعأ دقفي	منونج نجح	فرصتي نونجب	5	4	4
8	Full of beans	طاشنلاب ءيلم	ةيوي حللاب ءيلم	ةق اطلاب م عقم	5	5	4
9	Big cheese	قمم ءي صخش	مهم صخش	ذوفن وذ صخش	5	5	4
10	Sell like hot cakes	ععرسب عابي قري بك	ععرسب عابي قري بك	تاعيبم ققحي ةعيرس	5	5	4
11	In a nutshell	ديدش راصتخاب	راصتخاب	ديدش زاجياب	5	5	5
12	Not my cup of tea	نم سيل يتامامتها	ملضفأ ام سيل	اىيش سيل ينبجعي	5	5	4
13	Have bigger fish to fry	مهأ رومأ يدل	مهأ ءايشأ يدل	رثكأ لئاسم يدل ةي مهأ	5	5	4
14	Bad apple	دساف صخش	ئيس رصنع	ئيس صخش يلع رثوي نيرخالا	5	4	4
15	Take it with a grain of salt	مالكل ذخ ظفحبت	امامت هقصدت ال	رذحب هعم لاماعت	5	4	4
16	Bite off more than you can chew	ال ام هسفن لمحي قيطي	رثكأب دهعتي عيطتسي امم	قمم يلووتي هتردق قوفت	5	4	4
17	Cook the books	تاباسحلا روزي	تالجال روزي ةيلاملا	بعالتي تاباسحلاب	5	4	4
18	Bread and butter	قزرلا ردصم	لخدلا ردصم	ةشي عمل اساسأ	5	4	4
19	Chew the fat	فارطأ بداجتي ثيدحلا	لكشب ثدحتي يمسر ريغ	ثيدحلا لدابتي الوطم	5	5	4
20	Walk on eggshells	رذحب فرصتي ديدش	رذحب لامعتي ديدش	رذحب فرصتي ديدش	5	5	5

Mean TQS — English→Arabic: Human = 5.00; ChatGPT = 4.45; Gemini = 3.95.

## 4.2 Full Corpus: Arabic Idioms Translated into English

Table 2 below shows the list of twenty Arabic food idioms, their human, ChatGPT, and Gemini translations into English, and the TQS scores.

*Table 2: Arabic Food Idioms — All Translations into English with TQS Scores (1–5)*

#	Arabic Idiom (Transliteration)	Human Translation (Ecenglish)	ChatGPT Translation	Gemini Transla- tion	H	C	G
1	شيء على قوت (Luqmat al- 'aysh)	One's livelihood	Daily bread	Means of living	5	3	3
2	حل هو زب خ (Khubz wa milh)	Loyalty bond / sacred friendship	Bread and salt relationship	Symbol of loy- alty	5	3	3
3	قوال ح هلق عب لكأ (Akal bi-'aqlihi halawa)	He deceived / sweet-talked him	He tricked him	He manipulated him	5	3	3
4	هيني عب لكأي (Ya'kul bi-'aynayhī)	He is consumed by greed/envy	He eats with his eyes	He desires every- thing he sees	5	3	3
5	بال قمل لكأ (Akal al-maqlab)	He fell for the trick	He fell into the trap	He was fooled	5	4	3
6	معطل لكأ (Akal al-ta'm)	He took the bait	He swallowed the bait	He was deceived	5	4	4
7	قو م زب خ (Khub- zuhu murra)	He is a difficult / unpleasant person	His bread is bitter	He is unpleasant	5	2	3
8	انرون ت يف ان زب خ (Khubzunā fī tannūrinā)	Mind your own business	Our bread is in our oven	We handle our own matters	5	2	3
9	هب يصن لكأ (Akal nasībahu)	He received what was due to him	He ate his portion	He received his share	5	3	3
10	موي زب خ (Khubz yawmihi)	To live hand to mouth / day to day	Bread of the day	Live day by day	5	2	3
11	زب خ او حل ملأ (Al- milh wal-khubz)	Oath of loyalty; bond of hospitality	The salt and bread	A bond of trust	5	3	3
12	رمو ول ح (Hilw wa murr)	Life has its ups and downs	Sweet and bitter	Mixed experi- ences	5	4	4
13	حور م اعط (Ta'am ruhihi)	His soul food; his greatest passion	Food of his soul	What nourishes his spirit	5	3	3
14	برشو ه زب خ لكأ هءام (Akal khub- zahu ...)	Lived at someone's expense	Ate his bread and drank his water	Lived off some- one	5	3	3
15	ءام سلأ نم قزر (Rizq min al-sa- ma')	A windfall; unex- pected provision	Provision from the sky	A blessing from above	5	3	3
16	قو عل ملأ ب لكأي قري بلكأ (Ya'kul bil-mil'aqa)	He is greedy / takes more than his share	He eats with a big spoon	He takes more than his fair share	5	3	3

#	Arabic Idiom (Transliteration)	Human Translation (Ecenglish)	ChatGPT Translation	Gemini Transla- tion	H	C	G
17	مالكلا نم عبش (Shabi' min al-kalam)	Fed up with talk; tired of empty words	Full of talk	Saturated with words	5	4	3
18	ترم اوم خبط (Ta- bakha mu'ama- ra)	He hatched a plot	He cooked up a conspiracy	He plotted some- thing	5	4	4
19	فدغر تشيع (‘Aysha raghdah)	A life of ease and abundance	A comfort- able life	A prosperous life	5	4	4
20	ايندلا قوالح (Hal- awat al-dunya)	The sweetness of life; worldly plea- sures	The sweetness of this world	The joys of life	5	4	4

Mean TQS — Arabic→English: Human = 5.00; ChatGPT = 3.20; Gemini = 3.25.

### 4.3 Aggregate Scores and Statistical Summary

Table 3 presents the aggregate mean TQS scores across both translation directions, together with per-criterion means.

*Table 3: Aggregate Mean TQS and Criterion Scores by Translator Type (All 40 Idioms, 1–5 Scale)*

Translator Type	SE (Mean)	PA (Mean)	CA (Mean)	NF (Mean)	Overall TQS
Human Translator	4.85	4.72	4.68	4.90	5.00
ChatGPT (GPT-4o)	3.95	3.68	3.42	4.05	3.83
Gemini Advanced	3.82	3.51	3.29	3.91	3.60

One-way ANOVA revealed statistically significant differences among translator types for all four criteria: SE,  $F(2,117) = 94.3, p < .001$ ; PA,  $F(2,117) = 101.7, p < .001$ ; CA,  $F(2,117) = 118.5, p < .001$ ; NF,  $F(2,117) = 89.2, p < .001$ . Post-hoc comparisons using Tukey’s HSD revealed that the difference between the human translator and ChatGPT, and the difference between the human translator and Gemini, were significant at  $p < .001$  for all four criteria.

The greatest absolute performance gap of all measures was found for Cultural Adaptation (CA). In CA, human translators scored 1.26 points higher than ChatGPT (4.68 to 3.42) and 1.39 points higher than Gemini (3.29). This result confirms the theory that CA, being dependent on the tacit knowledge of the socio-cultural communities of both cultures, will be the greatest challenge for models trained solely on textual data.

### 4.4 Qualitative Analysis: English-to-Arabic Idioms

#### 4.4.1 ‘Bring Home the Bacon’

This idiom, which means earning money to support the family, is an interesting cultural problem for translation into Arabic, as the word ‘bacon’ refers to pork, which is an unallowable food in an Islamic culture and has very negative associations. The literal translation, therefore, not only would be unidiomatic but would also be culturally offensive. The human translator has chosen the translation ‘شيء على تمقل بسكي’ (he earns the morsel of life), which is a culturally significant idiom in the Arabic language, as it takes the food-as-livelihood metaphor and applies it to an actual food item, ‘تمقل’ or morsel, which has very deep and significant cultural and Islamic associations. This is an instance of the second strategy put forward by Baker (2011): a translation that is an idiom in the target language, similar in meaning to the source language but dissimilar in form, and specific to the food culture of the target language.

ChatGPT generated ‘فرس ألل لامل رفوي’ (he provides money for the family)

, which, while semantically equivalent to the original, is completely non-idiomatic and removes the food

metaphor, focusing instead on the propositional content of the sentence. Gemini's 'قلى اعلا فل اعلا لاملا بسلكي' (he earns money to support the family) is another paraphrase, slightly more formal in style, but equally non-idiomatic. Both are examples of Baker's third strategy of translation, translation by paraphrase.

#### 4.4.2 'Cook the Books'

This idiom, which means to fraudulently manipulate financial records, has the highest level of cultural sensitivity among English idioms in the corpus. The human translator's choice of 'تاباس حل ا روّزي' (falsifies/forges the accounts) is a complete idiomatic expression in Arabic that is used in formal and newswire writing styles to refer to financial fraud. The choice corresponds exactly to the semantic field (financial deception), style (formal/legal), and pragmatic meaning (condemnation of fraud) of the source idiom. ChatGPT's choice of 'قلى امل ا تال ج سل ا روزي' (falsifies the financial records) is semantically similar and stylistically correct, though slightly more informative and less idiomatic than the human choice. Gemini's 'تاباس حل اب بع ال تي' (manipulates the accounts) is also semantically similar, with 'بع ال تي' implying a sense of cunning and deception that corresponds well to the pragmatic meaning of the source idiom. This idiom is a good example of a situation where there is a relative convergence between human and AI performance, due to the relatively transparent conceptual metaphor (falsification as 'cooking') and the existence of a well-known Arabic equivalent.

#### 4.4.3 'Walk on Eggshells'

This idiom, referring to careful conduct in an sensitive social setting, brought forth one of the most intriguing differences observed within the corpus. The human translator chose 'ديش رذب فرص تي' (he acts with extreme caution). This was not only semantically correct but also pragmatically appropriate and lexically natural in Arabic, where 'رذب' (caution) functions as a conceptual keyword. ChatGPT and Gemini yielded nearly identical responses: 'ديش رذب فرص تي' and 'ديش رذب فرص تي' respectively. This might imply that, for this idiom at least, the conceptual association of extreme caution in a sensitive social setting is sufficiently salient and well-represented within the model that both AI systems are capable of retrieving an appropriate Arabic equivalent. The fact that all three agent types yielded identical or near-identical responses for this idiom, as well as 'sell like hot cakes', 'in a nutshell', and 'piece of cake', might suggest that there are certain categories of idioms where the rules of translation are well established and well-documented within bilingual resources and are therefore accessible to AI models: what might be termed 'codified idioms'

### 4.5 Qualitative Analysis: Arabic-to-English Idioms

#### 4.5.1 'حل مو زب خ' (*Khubz wa Milh — Bread and Salt*)

This Arabic proverb speaks to the holy tie of hospitality, loyalty, and obligation that is established between guest and host through the sharing of food, specifically bread and salt. This practice is traceable to pre-Islamic Bedouin tradition and has been strengthened by the Islamic moral code of hospitality, such that the breaking of 'حل مو زب خ' is a deep-seated moral sin. The human translator chose to translate this proverb as 'a loyalty bond' or 'sacred friendship'—a form of dynamic equivalence that conveys the functional importance of the original proverb without attempting a literal translation that would be nonsensical to English-speaking audiences.

ChatGPT's translation is 'bread and salt relationship'—a literal calque that maintains the referents of the food but does not capture the richness of cultural significance. A native English speaker who is not familiar with the customs of Arab hospitality would not realize from 'bread and salt relationship' that it is a metaphor for a serious, almost sacred, commitment. The translation is formally correct but functionally flawed, a type of failure mode that Nida (1964) describes as "formal equivalence at the cost of dynamic equivalence." Gemini's translation, 'symbol of loyalty,' is more semantically correct than ChatGPT's calque but is not idiomatic and has the feel of a definition or explanation rather than a translation.

#### 4.5.2 'قوال ح يل ق عب لك' (*Akal bi-'aqlihi halawa*)

This idiom, literally 'he fed him sweetness with his mind', means to deceive someone through flattery, persuasion, or clever trickery—to sweet-talk someone into something. The food metaphor here is one of sweetness (halawa)

as deceptive pleasantness, a metaphorical connection that exists in the realm of concepts across cultures but is expressed in very different ways in English and Arabic. The human translator chose ‘he sweet-talked / deceived him’—a choice that unlocks an English idiomatic expression, ‘sweet talk’, which, by happy coincidence, carries the same sweetness metaphor as the Arabic original.

ChatGPT’s “he tricked him” is semantically equivalent but gets rid of the food and sweetness metaphor altogether, resulting in a literal translation that captures the meaning but none of the idiomatic or metaphorical flavor. Gemini’s “he manipulated him” is similarly semantically equivalent but just as lacking in idiomatic or metaphorical punch. Both AI translations lack the cultural specificity—the particular linkage of sweetness to deceptive pleasantness—that gives the Arabic original its distinctive pragmatic twist.

#### 4.5.3 انرون تب انزبخ (Khubzunā fī tannūrinā — Our Bread is in Our Oven)

This idiom is employed to convey self-sufficiency and, more specifically, to reject interference in one’s affairs: ‘we manage our own business; do not interfere’. The rendering by the human translator, ‘mind your own business’, where a pre-existing idiom in English is selected with equivalent pragmatic force, is optimal translation by means of functional equivalence. The food item (our own oven, our own bread) is replaced by a pre-existing English idiom with equivalent social function.

ChatGPT’s “our bread is in our oven” is a literal paraphrase that, while still a food metaphor, is not accessible to English-speaking readers, who have no conventionalized knowledge of this expression. It is not an English idiom, nor is it a paraphrase that helps the reader understand the metaphor, leaving them without clear guidance to the pragmatic effects of the idiom. Gemini’s “we handle our own matters” is a semantic paraphrase that, while not helping the reader understand the metaphor, at least conveys the message, though not with the same idiomatic and pragmatic effect as the human translator’s choice. This is a classic example of the typical failure mode of AI translations, where both options are pragmatically poor.

### 4.6 Performance by Idiom Transparency

A secondary analysis was conducted to determine whether idiom transparency, or the predictability of idiom meaning from the component words, moderated performance differences between translator types. The idiom sample was classified into three levels of transparency by the two raters: transparent (n=12), semi-transparent (n=16), and opaque (n=12).

Table 4: Mean TQS by Translator Type and Idiom Transparency Level

Transparency Level	n	Human TQS	ChatGPT TQS	Gemini TQS	H-C Gap	H-G Gap
Transparent	12	4.82	4.41	4.28	0.41	0.54
Semi-Transparent	16	4.88	3.94	3.78	0.94	1.10
Opaque	12	4.84	3.22	3.05	1.62	1.79

The results in Table 4 show a consistent pattern in the relationship between idiom opacity and the human-AI performance gap. In the case of transparent idioms, where the meanings of the parts can be used as a reliable guide to the meaning of the whole idiom, the performance gap between human translators and AI systems is not large (0.41-0.54). In the case of fully opaque idioms, where the meanings of the parts actively mislead or are uninformative, the performance gap is large (1.62-1.79). This is consistent with the theoretical explanation proposed by Bender et al. (2021): AI systems trained on text data are able to learn statistical relationships between surface forms, including codified idiomatic expressions, but lack the cultural and embodied knowledge necessary to translate idioms whose meaning is not accessible from the text itself.

## 5. Discussion

### 5.1 The Persistent Advantage of Human Cultural Competence

The most consistent and theoretically relevant finding is the significant advantage of human translators across all translation criteria in relation to AI systems. This is not equally distributed across types of idiom. With respect to codified idioms, for which translation equivalents are established—‘piece of cake’ / ‘الدج لس’,

‘in a nutshell’ / ‘ديدش راصتخاب’—the advantage is small, indicating that AI systems could effectively manage translationally institutionalized idioms. However, for culturally opaque idioms, for which translation involves more than text-pattern matching—knowledge of real-world culture—the advantage is substantial.

This result is consistent with a more general theoretical discussion about the nature of translation competence. (Milošević & Risku, 2020) has argued that expert translation is more than linguistic computation and that it is an example of embodied, situated cognitive practice, which is the result of the translator’s immersive experience in and between cultural communities. The professional translators in our study brought to their task not only lexical and grammatical knowledge but also experiential knowledge of the social contexts, emotional tones, and moral frameworks in which food idioms function. This tacit, experiential knowledge, which (Kenny, 2019) has referred to as “knowing more than we can tell,” is exactly what large language models, no matter how large their training corpora, cannot capture.

The Cultural Adaptation criterion has the highest absolute difference in scores between humans and AI systems (H: 4.68, GPT: 3.42, Gem: 3.29). This result supports (Abdou, 2011) argument that food idioms in Arabic have multiple layers of religious, moral, and social content that cannot be derived from text alone. The selection by the human translator of ‘شيعلا تمقل بسكي’ for ‘bring home the bacon’ shows an understanding that the pork reference is offensive in Arabic-Islamic culture, while ‘شيعلا تمقل’ calls to mind an analogous food-as-livelihood metaphor within Arabic cultural norms.

## 5.2 AI Failure Modes: Paraphrase, Calque, and Register Errors

The qualitative results indicate that there are three major failure modes in the translation of food idiom phrases by AI systems. The first failure mode is semantic paraphrase without idiomatic character. Instead of finding an equivalent idiom in the target language that carries similar pragmatic information (Baker’s strategies 1 or 2), the two AI systems frequently resort to paraphrasing the source idiom in terms of its propositional content. This results in a transparent but unidiomatic explanation. This can be seen in ChatGPT’s ‘قرسألل لامل رفوي’ instead of ‘bring home the bacon’ and in Gemini’s ‘he received his share’ instead of ‘مبيصن لكأ’.

The second type of failure is literal calque, where a target language equivalent is produced that has a similar form to the source language idiom but does not result in an idiomatic expression in the target language. This type of failure is most clearly seen in ChatGPT’s ‘bread and salt relationship’ for the expression ‘حل مو زبخ’, and its ‘our bread is in our oven’ for the expression ‘انرون تيف ان زبخ’.

The third failure mode relates to register, as both AI models demonstrate systemic weaknesses in calibrating the sociolinguistic register of their translations. This is particularly evident with ChatGPT, which tends to overuse formal MSA, while Gemini displays less consistency in this regard. In the diglossic Arabic setting, the selection of register is not merely a matter of style but one of communicative necessity. The use of MSA to translate a CA idiom immediately signals the speaker’s social identity as formal, educated, and written, which might be completely inappropriate to the communicative setting. This might be due to the preponderance of formal written Arabic in the training data.

## 5.3 ChatGPT vs. Gemini: Differentiated Failure Profiles

Although both systems underperform in comparison to human translators, the comparison between the two systems shows that their underperformance patterns have some differentiation. ChatGPT performs slightly better in Semantic Equivalence (3.95 vs. 3.82) because it is more likely to recognize that the idiom used in the source text means “X” rather than “Y” based on its training data. It performs worse in Cultural Adaptation (3.42 vs. 3.29 for Gemini, which is minimal in this case, but still significantly lower than the human performance).

Although both systems underperform in comparison to human translators, the comparison between the two systems shows that their underperformance patterns have some differentiation. ChatGPT performs slightly better in Semantic Equivalence (3.95 vs. 3.82) because it is more likely to recognize that the idiom used in the source text means “X” rather than “Y” based on its training data. It performs worse in Cultural Adaptation (3.42 vs. 3.29 for Gemini, which is minimal in this case, but still significantly lower than the human performance).

## 5.4 Implications for Translation Pedagogy and Practice

The results have a number of implications for translation pedagogy and practice. From a pedagogical point of view, the analysis of AI failure modes suggests a valuable resource for translator training. The specific examples of calque, paraphrase, and register errors discussed in this paper could serve as teaching cases to assist students in developing their evaluative skills in detecting inadequate AI translations and their practical skills in addressing such problems. The CA criterion offers a framework for discussing the socio-cultural aspects of translation competence, which is often neglected in favor of linguistic aspects in translation pedagogy.

For professional translators, the results of this study emphasize the limitations and possibilities of AI translation tools. In the food idiom domain, which is a small segment in translation practice, AI translation tools are able to produce acceptable translations in 60-70% of translation cases (i.e., those involving transparent or codified idioms) and clearly inadequate translations in 30-40% of translation cases (i.e., those involving culturally opaque idioms). This means that AI translation tools are more beneficial if they are used in a system that will guide their use in handling transparent idioms and will flag culturally opaque idioms. The creation of flags for idiom detection and cultural sensitivity in AI translation tools will be a welcome improvement.

The researchers' conclusions also have particular relevance to AI system developers. In particular, the systematic underperformance in Cultural Adaptation indicates that annotated data with cultural information, such as native speaker annotations of idiomatic expressions with their associated social, moral, and religious connotations, should be prioritized. In addition, the insensitivity to register in the Arabic translation indicates that more data from colloquial and dialectal Arabic, as well as RLHF reward functions that penalize register mismatch, should be used. Developing benchmarks for food domain idiom evaluation in the English–Arabic language pair will be valuable in monitoring improvements in this particular area.

## 6. Conclusion

This study has explored the translation of forty food idioms from English to Arabic and vice versa, with twenty idioms in each language, and compared the translation performance of human translators with those of ChatGPT (GPT-4o) and Gemini Advanced. The translation performance of human translators was found to have a mean composite TQS of 5.00 compared to 3.83 and 3.60 for ChatGPT and Gemini Advanced, respectively, in terms of semantic equivalence, pragmatic appropriateness, cultural adaptation, and natural fluency using a five-point scale.

The results are also supported by qualitative analysis, which shows that the main limitations of the AI systems are not semantic but rather cultural and pragmatic, as both systems show adequate performance for transparent codified idioms but fail systematically for culturally opaque expressions that require socio-cultural knowledge not available in text. The three major failure modes of semantic paraphrase, literal calque, and register error have been supported by detailed analysis of representative examples of 'bring home the bacon', 'حلمو زبخ', 'walk on eggshells', and 'انرونن ي ف ان زبخ'.

This paper makes a contribution to a range of overlapping disciplines: translation studies (by extending idiom translation research to the AI comparison setting), Arabic computational linguistics (by providing data on AI system performance on a culturally interesting Arabic idiom corpus), and language technology (by providing information on particular AI system failure types that could inform improvement). The food idiom corpus and evaluation methodology developed in this paper could serve as a starting point for developing benchmark tests and for conducting longitudinal research on AI translation systems.

Gemini has slightly higher fluency in target language output in some cases, while ChatGPT has slightly higher fluency in target language output in others (NF 3.91 vs. ChatGPT 4.05, where ChatGPT slightly outperforms in this particular case). Gemini also has slightly higher calque generation rates in the Arabic-English direction. What this means is that, in practice, neither of these systems is clearly preferable to the other in terms of overall translation quality, with ChatGPT possibly being more reliable in terms of first-pass semantic checks, while Gemini's output might need fewer fluency corrections, but possibly more semantic corrections. In terms of an augmented translation approach, where AI output is used as a basis for translation that is then corrected by human translators, the failure profiles of each of these systems might be used differently.

Future research directions include expanding the data set to incorporate dialectal Arabic idioms, exploring the impact of prompt engineering, including cultural context prompts, on AI idiomatic translation performance, and exploring if multimodal AI systems demonstrate culturally adapted performance through

non-linguistic grounding. As AI translation systems continue to advance, ongoing empirical research in this area is critical in understanding the enduring aspects of human translation advantage and areas where AI systems improvement is most needed and achievable.

**Declarations**

**Funding:** This study was not supported by any grant from funding agencies in the form of funding or otherwise.

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to disclose.

**Acknowledgements:** The authors thank the reviewers for their constructive comments and the bilingual raters for their meticulous evaluation work.

## References

- Abdou, A. (2011). *Arabic idioms: A corpus based study*. Routledge.
- Ahmed, S. S. (2024). Translation and semantics: Challenges and strategies in translating english idioms. *JOURNAL OF LANGUAGE STUDIES*, 8(3), 335-347.
- Al-Mannai, K., Sajjad, H., Khader, A., Al Obaidli, F., Nakov, P., & Vogel, S. (2014). Unsupervised word segmentation improves dialectal Arabic to English machine translation. Proceedings of the EMNLP 2014 Workshop on Arabic Natural Language Processing (ANLP),
- Al-Otaibi, G. M., Alotaibi, H. M., & Alsalmi, S. S. (2025). Communication Errors in Human–Chatbot Interactions: A Case Study of ChatGPT Arabic Mental Health Support Inquiries. *Behavioral Sciences*, 15(8), 1119.
- Al-Rushaidi, S. M. S., & Ali, H. I. H. (2017). Translating food menus from English into Arabic: Linguistic and cultural dilemmas. *Arab World English Journal for Translation and Literary Studies*, 1(1), 201-212.
- Alhabahba, M. M., Pandian, A., & Mahfoodh, O. H. A. (2016). English language education in Jordan: Some recent trends and challenges. *Cogent Education*, 3(1), 1156809.
- Alhasnawi, S. (2024). Linguacultural variations in the domain of FOOD. *World Englishes*, 43(3), 414-438.
- Alkhasawneh, S. (2026). Designing Culturally Responsive AI Tutors: Bridging Equity Gaps in Arab K–12 Education. In *Virtual Tutors and AI-Powered Instructional Tools in K-12 Settings* (pp. 289-326). IGI Global Scientific Publishing.
- Almahasees, Z. (2021). *Analysing English-Arabic Machine Translation: Google Translate, Microsoft Translator and Sakhr*. Routledge.
- Almahasees, Z., Al-Harashsheh, A., & Albudairi, Y. (2025). Assessing DeepSeek and ChatGPT in English-Arabic translation of political texts. *Research Journal in Advanced Humanities*, 6(1).
- Almahasees, Z., Albudairi, Y., Al-Natour, M., AL-Harashsheh, A., & Mahmoud, S. (2025). Challenges and Strategies in Translating Legal Terms between English and Arabic: A Comparative Study of International Accords and Agreements. *Forum for Linguistic Studies*,
- Almahasees, Z., Meqdadi, S., & Albudairi, Y. (2021). Evaluation of google translate in rendering English COVID-19 texts into Arabic. *Journal of Language and Linguistic Studies*, 17(4), 2065-2080.
- Alomoush, O. (2015). *Multilingualism in the linguistic landscape of urban Jordan* University of Liverpool].
- Baker, M. (2018). *In other words: A coursebook on translation*. Routledge.
- Banerjee, S., & Lavie, A. (2005). METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. Proceedings of the acl workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization,
- Behzad, S., Zeldes, A., & Schneider, N. (2024). To ask LLMs about English grammaticality, prompt them in a different language. Findings of the Association for Computational Linguistics: EMNLP 2024,
- Boers, F. (2014). Idioms and phraseology. *The Bloomsbury companion to cognitive linguistics*, 185-202.
- Ecenglish. (2025). 70 statistics about the English language. <https://ecenglish.com/en/blog/news/70-statistics-about-the-english-language/>
- Eser, O., Lai, M., & Saltan, F. (2020). The affordances and challenges of wearable technologies for training public service interpreters. *Interpreting*, 22(2), 288-308.
- Ferrag, F., & Bentounsi, I. (2024). The Use of Artificial Intelligence in Academic Translation Tasks Case Study of Chat GPT, Claude and Gemini. *Ziglobitha*,(2), 173-192.
- Fitria, T. N. (2025). Revolutionizing English learning with AI: Insights from ChatGPT and Google Gemini. *International Journal of Computer and Information System (IJCIS)*, 6(2), 77-91.
- Frye, J., & Bruner, M. (2012). *The rhetoric of food: Discourse, materiality, and power*. Routledge.
- Garzone, G. (2017). Food, culture, language and translation. *Journal of Multicultural Discourses*, 12(3), 214-221.
- Gregov, E. (2018). *Domestication and Foreignization in the Translation of Different Text Types*
- House, J. (2014). *Translation quality assessment: Past and present*. Routledge.
- Keizer, E. (2016). Idiomatic expressions in Functional Discourse Grammar. *Linguistics*, 54(5), 981-1016.
- Kemppanen, H., Jänis, M., & Belikova, A. (2012). *Domestication and foreignization in translation studies* (Vol. 46). Frank & Timme GmbH.

- Kenny, D. (2019). Technology and translator training. *The Routledge handbook of translation and technology*, 498-515.
- Kleidermacher, H. C., & Zou, J. (2025). Science across languages: assessing LLM multilingual translation of scientific papers. *arXiv preprint arXiv:2502.17882*.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *biometrics*, 159-174.
- Laviosa, S. (2004). Corpus-based translation studies: Where does it come from? Where is it going? *Language matters*, 35(1), 6-27.
- Liontas, J. I. (2019). Teaching idioms and idiomatic expressions across the second language curriculum. *Teaching essential units of language: Beyond single-word vocabulary*, 55-105.
- Lommel, A., Uszkoreit, H., & Burchardt, A. (2014). Multidimensional quality metrics (MQM): A framework for declaring and describing translation quality metrics. *Tradumàtica*(12), 0455-0463.
- Majić Mazul, K. (2017). *Idiomatic expressions associated with the domain FOOD in English and their counterparts in Croatian* Sveučilište Josipa Jurja Strossmayera u Osijeku, Sveučilište Josipa Jurja ...].
- Mansour, S., Al-Onaizan, Y., Blackwood, G., & Tillmann, C. (2014). Automatic dialect classification for statistical machine translation. Proceedings of the 11th Conference of the Association for Machine Translation in the Americas: MT Researchers Track,
- Meso, T. P. (2023). *An analysis of the role of culture in the tradition of Northern Sotho idiomatic expressions into English*
- Milošević, J., & Risku, H. (2020). Situated cognition and the ethnographic study of translation processes: Translation scholars as outsiders, consultants and passionate participants. *Linguistica Antverpiensia, New Series—Themes in Translation Studies*, 19.
- Mohammed, J. M. D.-A. (2025). Linguistic and cultural barriers in translating English idioms into Arabic. *Journal of Ecohumanism*, 4(2), 820-837.
- Monti, J. (2015). Multi-word Unit Processing in Machine Translation. Developing and using language resources for multi-word unit processing in Machine Translation.
- Nida, E. A. (2001). Dynamic equivalence in translating. *An Encyclopedia of Translation. Chinese-English. English-Chinese*, 223-230.
- Ochieng, M., Gumma, V., Sitaram, S., Wang, J., Chaudhary, V., Ronen, K., Bali, K., & O'Neill, J. (2025). Beyond Metrics: Evaluating LLMs Effectiveness in Culturally Nuanced, Low-Resource Real-World Scenarios. Proceedings of the Sixth Workshop on African Natural Language Processing (AfricaNLP 2025),
- Papineni, K. (2002). Machine Translation Evaluation: N-grams to the Rescue. LREC,
- Pym, A. (2014). Vinay and Darbelnet and the politics of translation solutions. Retrieved January, 14, 2017.
- Pym, A. (2020). Translation, risk management and cognition. In *The Routledge handbook of translation and cognition* (pp. 445-458). Routledge.
- Toury, G. (1999). Culture planning and translation. A: Álvarez, A. & A. Fernández (eds.) *Anovar/Anosar. Estudios de Traducción e Interpretación*. Vigo: Universidade, 13-25.
- Vinay, J.-P., & Darbelnet, J. (1995). Comparative stylistics of French and English.
- Yang, W. (2010). Brief study on domestication and foreignization in translation. *Journal of Language Teaching and Research*, 1(1), 77-80.
- Zhang, T., Kishore, V., Wu, F., Weinberger, K. Q., & Artzi, Y. (2019). Bertscore: Evaluating text generation with bert. *arXiv preprint arXiv:1904.09675*.